THE UNIVERSITY
of EDINBURGH

Characterising gene-gene and gene-environment interactions within the fitness landscape of a snoRNA

Luke Jonathan Williamson

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that this thesis has been composed by myself, and is a result of my own work unless explicitly stated otherwise. It has not previously been submitted for any other type of degree, diploma or qualification.

_____

Luke Jonathan Williamson

Date: _____

Abstract

Empirical fitness landscapes allow a previously hypothetical concept describing the genotype-phenotype relationship to be visualised and dissected. The fitness landscapes of entire molecules have only recently begun to be elucidated, with gene-gene and gene-environment interactions being largely unexplored in this context. U3 small nucleolar RNA (snoRNA) represents a suitable model in which to examine these interactions, with the strategy of deep mutational scanning enabling mutational effects to be studied at a single nucleotide resolution. In yeast, the U3 snoRNA directs three essential cleavage events of the pre-ribosomal RNA (pre-rRNA), facilitated by complementarity between the 5' region of U3 snoRNA and the 35S pre-rRNA transcript. These cleavage events along with folding of the pre-rRNA are facilitated by a series of proteins, some of which bind with U3 snoRNA to form the U3 snoRNP.

Since U3 snoRNA interacts with several proteins, it is an ideal candidate to study the molecular mechanisms behind gene-gene interactions. I have used the auxin inducible degron system to downregulate eight proteins that interact with U3 snoRNA. The disruption of these essential genes demonstrated intermolecular epistasis, altering the fitness landscape in a gene specific manner. Conditionally deleterious mutations were enriched in regions of U3 snoRNA whose role is related to that of the hypomorphic protein.

To simulate the changing environments that an organism may face, and observe the resulting impact, I measured the fitness landscape of U3 snoRNA in three different temperature environments. This highlighted numerous gene-environment interactions, related to structural and protein binding regions of U3 snoRNA. Various genotypes tolerated at physiological temperature become deleterious at higher and lower temperatures.

Lay Summary

The relationship between the genes an organism may possess and the resulting consequence is of great interest. Knowing how each variant of a gene may act provides a predictive power useful both in biological and medical context. This relationship is also central to understanding evolution, as changes in the gene occur first, with selection then acting upon the consequence of said change. By setting up experiments to simulate and scrutinise this connection, we can effectively quantify selection acting upon an organism. In this PhD project I aimed to study the effect of thousands of potential changes (mutations) to a gene, using computational methods to then accurately measure the negative, positive or neutral effect of each one. I examined whether these effects were different in different conditions. I performed these experiments in yeast, an organism that shares a lot of biology with humans, and looked at mutations in a gene important to making new proteins.

To simulate the changing environments that an organism may face, I looked at the effect of mutations in temperatures colder and warmer than usual. This exposed mutations that while tolerated at normal temperatures, were now having large negative effects at different temperatures. This shows how evolution shapes an organism to be suited to a particular environment.

I then manipulated proteins that work with my gene of interest to be at lower levels, to see how the mutations in the gene and the change in protein level interact. I found that the resulting changes were very specific to which protein was disrupted, with mutational effects in areas related to the proteins role becoming exacerbated. This showed how interconnected interacting molecules are.

Abbreviations

Acronym	Meaning
%	Percentage
°C	Celsius degreees
A	Adenine
AID	Auxin inducible degron
C	Cytosine
cDNA	Complementary DNA
Cryo-EM	Cryoelectron microscopy
CSM	Complete supplement mixture
DAmP	Decreased abundance by mRNA perturbation
ddH_20	double distilled H_20
DFC	Dense fibrillary component
Dhr1	DEAH-Box Helicase 1
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
ETS	External transcribed spacer
FC	Fibrillar centre
G	Guanine
GC	Granular component
GFP	Green fluorescent protein
IAA	Indole-3-acetic acid
ITS	Internal transcribed spacer
KPi	Potassium phosphate buffer
k-turn	Kink-turn
LiAc	Lithium acetate
LSU	Large subunit (ribosome)
M	Molar
mg	Milligrams
min	Minutes
mL	Millilitres
mM	Millimolar
Mpp10	M-Phase Phosphoprotein 10
mRNA	Messenger RNA
NAA	1-naphthaleneacetic acid
NaCl	Sodium chloride

ng	Nanograms
Nop1	Nucleolar Protein 1
Nop56	Nucleolar Protein 56
Nop58	Nucleolar Protein 58
OD_{600}	Absorbance at wavelength of 600 nM
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pre-mRNA	Precursor mRNA
pre-rRNA	Precursor rRNA
qRT-PCR	Quantitative RT followed by PCR
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
r-protein	Ribosomal protein
rRNA	Ribosomal RNA
Rrp9	Ribosomal RNA Processing 9
RT	Reverse transcription
S. cerevisiae	*Saccharomyces cerevisiae*
S. kudriavzevii	*Saccharomyces kudriavzevii*
SDS	Sodium dodecyl sulfate
snoRNA	small nucleolar RNA
snoRNP	Small nucleolar ribonucleoproteins
Snu13	Small Nuclear Ribonucleoprotein 13
SSU	Small subunit (ribosome)
T	Thymine
tRNA	Transfer RNA
UTR	Untranslated region
Yeast	*Saccharomyces cerevisiae*
YMM	Yeast minimal media
YNB	Yeast nitrogen base

Table of Contents

Table of figures

5 INVESTIGATING GENE-ENVIRONMENT INTERACTIONS WITHIN THE FITNESS LANDSCAPE OF U3 SNORNA

1 INTRODUCTION

1.1 The U3 small nucleolar RNA

1.1.1 Ribosome biogenesis

Ribosomes are molecular machines comprised of RNA and protein, carrying out the essential function of protein synthesis. Due to the importance of protein synthesis to the survival and proliferation of cells, ribosomes are extremely abundant in the cell, with ~83% of the total cellular RNA being ribosomal (von der Haar, 2008). In a rapidly growing yeast cell this equates to 60% of cellular transcription, with ~2000 ribosomes assembled every minute (Warner, 1999).

In yeast more than 200 nonribosomal proteins along with RNA trans-acting factors have been implicated in ribosome biogenesis, a metabolically-expensive multi stage process (Woolford & Baserga, 2013). It involves the assembly of four rRNAs and around 79 r-proteins into a large 60S and small 40S subunit. The small 40S ribosomal subunit containing 18S rRNA and 33 proteins decodes mRNA via tRNAs, while the large 60S subunit containing 5S, 5.8S, 28S rRNA and 46 proteins catalyses peptide-bond formation via the peptidyl transferase reaction. The 5S rRNA is transcribed by RNA polymerase III, while 5.8S, 18S and 28S rRNAs are cleaved from a 35S pre-rRNA transcribed by RNA polymerase I.

The synthesis, processing and assembly of ribosomes occurs in the nucleolus, a sub-nuclear structure in the cell devoted to this purpose. In yeast the non-membrane bound region is defined by a tandem array of ~150 rDNA repeats on chromosome XII (Thiry & Lafontaine, 2005). The 35S transcript is transcribed from these rDNA repeats, with ribosome assembly factors binding to begin the multifaceted process required for maturation. The sequential addition of assembly factors leads first to the formation of the ribosomal subunit

precursor termed the SSU processome, and then formation of the LSU processome. Cleavage of the transcript at ITS1 separates the 35S transcript and requisite rRNAs, facilitating progression of the subunit precursors along their respective lineages. A subsequent highly coordinated series of cleavages, modifications, chaperoning and incorporation of ribosomal proteins leads to the independent maturation of each subunit. This process transitions first from the nucleolus to the nucleoplasm, and then to the cytoplasm to become mature ribosomes.

1.1.2 Small subunit processome formation

The nucleolus can be subdivided into three compartments, the fibrillary centre (FC), dense fibrillary component (DFC) and granular component (GC). The rDNA is situated in the FC, with transcription occurring into the DFC and subsequent cleavage occurring within in the DFC. The final nucleolar ribosome biogenesis steps then occur in the GC. Transcription of the pre-rRNA 35S transcript first occurs from the rDNA repeats, which are not all functionally active at once, but are thought to alternate (Tan & van Oudenaarden, 2010). The abundance of inactive repeats helps maintain the vital concerted evolution of these areas, as highly transcribed regions are unable to establish sister-chromatid cohesion required to facilitate recombinational repair (Ide, et al., 2010).

The 35S transcript requires significant modification, cleavage and folding before becoming functionally active within ribosomes. The earliest stages of pre-rRNA processing involves the assembly of the SSU processome on the 35S transcript. This early stable intermediate was first observed on Miller spreads using electron microscopy, seen as terminal knobs on 5’ end of nascent transcripts from the rDNA locus (Miller & Beatty, 1969). U3 snoRNA

was later purified with 28 proteins that were identified as the terminal knob particle found on the 35S transcripts and termed the SSU processome (Dragon, et al., 2002). The SSU processome is a RNP which forms cotranscriptionally (Osheim, et al., 2004), containing the pre-rRNA, a series of ribosome assembly factors, r-proteins and the U3 snoRNP (Grandi, et al., 2002). There is an involvement of ~100 assembly factors with a mass of ~6MDa, being 4-fold larger than the mature SSU. Recently a much more complete view of the yeast SSU processome has been obtained by cryo-EM (Barandun, et al., 2017), providing positional information and complimenting biochemical findings.

The hierarchical assembly of the SSU processome is coordinated with the transcription of the 35S transcript (**Figure 1**). As the 5' ETS, the domains of the pre-18S and ITS1 are transcribed, ribosome assembly factors are cotranscriptionally added to create an interconnected scaffold (Zhang, et al., 2016), (Chaker-Margot, et al., 2015 , Kornprobst, et al., 2016). This begins with the UtpA complex (Utp4, Utp5, Utp8, Utp9, Utp10, Utp15, Utp17), UtpB complex (Utp1, Utp6, Utp12, Utp14, Utp18, Utp21) and U3 snoRNP (U3 snoRNA, Nop1, Snu13, Nop56, Nop58, Rrp9), being assembled as the 5' ETS is transcribed. These are the major complexes that comprise the 2MDa 5'-ETS particle, the structural base of the SSU processome (Pérez-Fernández, et al., 2007 , Hunziker, et al., 2016). U3 snoRNA takes advantage of its complimentary to regions of the pre-rRNA to direct specific processing events. UtpA is important in stabilising the spacer RNA helices, and UtpB stabilises the interactions between U3 snoRNA and the 5' ETS (Barandun, et al., 2017). In the timeline of assembly, three subunits of UtpA (Utp8, Utp9, Utp17) bind first to the 5' ETS, while the remaining subunits of UtpA interact with nucleotides downstream. UtpB first binds via its Utp18 subunit, with Utp1 then binding to the duplex between U3 snoRNA and the 5' ETS. This base pairing

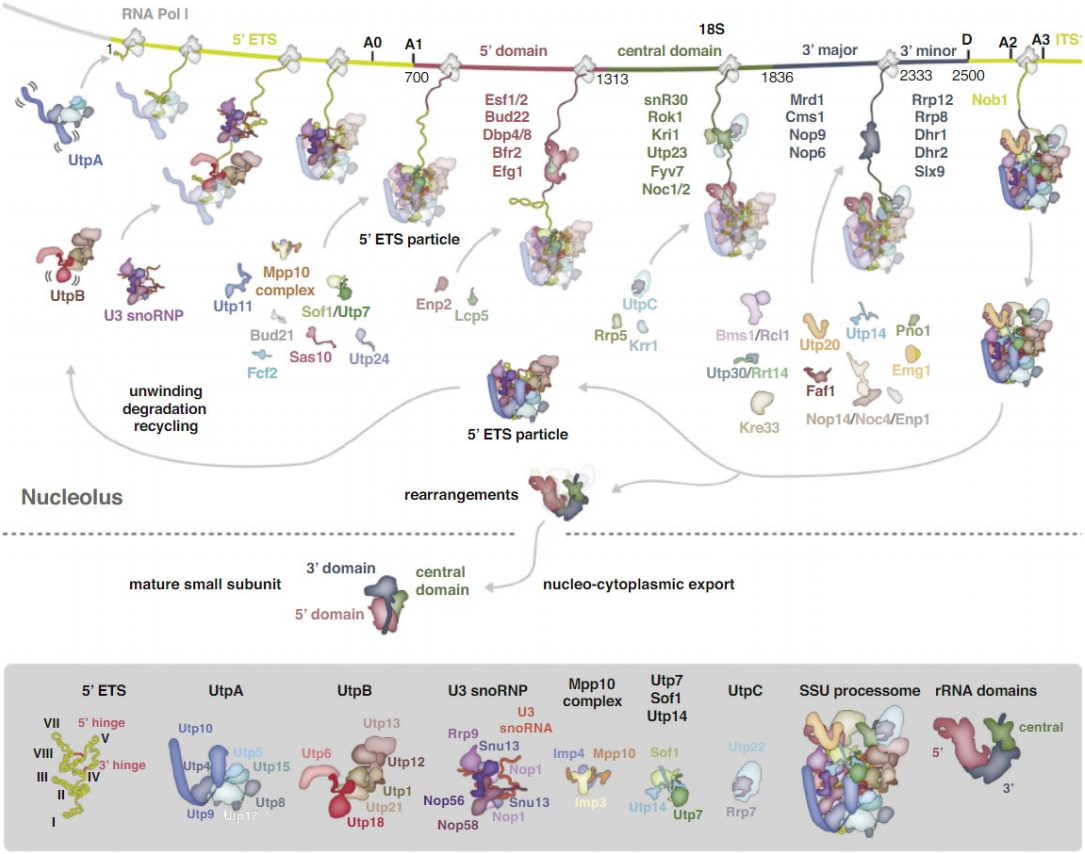


Figure 1, Cotranscriptional assembly of the SSU processome

Schematic representation showing the transcription of the 35S pre-rRNA transcript, and the successive cotranscriptional association of ribosome assembly factors. The different stages of assembly are illustrated as progression occurs along the different regions of the pre-rRNA. The different particles are labelled, with the complexes labelled in the lower grey box. (From (Barandun, et al., 2018)).

between U3 and the 5’ ETS helps stall progression of the central pseudoknot, a tertiary structure of the 18S that will determine rRNA domains in the mature ribosome. The pre-18S region of the transcript can be classified into four domains (5’ domain, central domain, 3’ major, 3’ minor), with different factors binding as successive domains are transcribed (Chaker-Margot, et al., 2015). This begins in the 5’ domain as the Mpp10 complex (Imp3, Imp4, Mpp10) along with a collection of other factors bind to complete the 5’-ETS particle. This includes factors such as Utp24 (Wells, et al., 2016) and Rcl1 (Horn, et al., 2011), nucleases responsible for early cleavage events that diverge the separate subunit pathways, essential for SSU processome maturation. Following on from the 5’-ETS particle the subsequent factors constituting the SSU are assembled.

After these waves of assembly factors and r-proteins is finished, and the ITS1 is transcribed, the SSU processome is complete. Changes then begin for maturation into the pre-40S particle. The major steps in the maturation of the SSU processome include the removal of the initial 5’ETS particle, folding of the central pseudoknot and cleavage of the pre-rRNA. The U3 snoRNA whose base pairing was structurally vital, must be released, thought to partially unwound by the Dhr1 helicase (Sardana, et al., 2015) and Mpp10 complex proteins Imp3 and Imp4 (Shah, et al., 2013). This unwinding is thought to also be facilitated by the UtpB subunit Utp14, which initially obscures the duplex, and so requires conformational change for it to be exposed. This conformation may be a way to communicate the SSU processome assembly status to both Dhr1 and the exosome (Black, et al., 2018). This SSU processome conformation with the site protected is identical in cryo-EM from Dhr1 depleted cells (Sun, et al., 2017). After these requisite steps occur, assembly factors are rearranged along with large compaction of the RNA transitioning into the pre-40S particle. After this many factors are recycled, forming a 5’ ETS recycling

intermediate with superfluous RNA being degraded by the exosome (Kornprobst, et al., 2016).

A recent Cryo-EM map of the SSU processome (Barandun, et al., 2017) has allowed validation and rationalisation of many previous findings. At 3-3.8 Å this structure of the SSU processome preceding the cleavage at site A_1 is able to show 15 r- proteins, 51 assembly factors, and U3 snoRNA bound to the pre-rRNA. The SSU processome is a globular particle, with the 5' ETS particle at the base, and U3 snoRNA facilitating base pairing with the pre-rRNA. From this 5' ETS particle other proteins expand out to interact with other factors and pre-rRNA. The SSU processome appears to function as a scaffold to postpone maturation by keeping pre-rRNA domains separate until vital processing and modification steps are complete. Factors within the SSU processome can be characterised as having an RNA protection/stabilisation role, a bridging role or as enzymes. (Chaker-Margot, 2018). The pre-rRNA is protected by preventing degradation and stopping the formation of incorrect structures. For example Mpp10 stabilises and unfolds helix 44 of the pre-18S, distancing it from certain pre-rRNA domains and postponing subsequent tertiary interactions that form the central pseudoknot. Similarly factors Bms1, Nop14, Faf1 and Imp4 stop premature pseudoknot formation by stabilising the association between U3 snoRNA and domains that will become the central pseudoknot in the final conformation.

1.1.2.1 Small nucleolar RNAs and RNA modification

The 35S transcript is substantially modified during processing, which can occur cotranscriptionally (Kos & Tollervey, 2010). While the disruption of individual modifications can be tolerated, removing them completely significantly effects growth (Tollervey, et al., 1993 , Zebarjadian, et al., 1999), indicating that

collectively they work in a synergistic way for proper ribosome biogenesis. This reduction in ribosome function is thought to stem from a reduction in translation efficiency (Baxter-Roshek, et al., 2007). These modifications come in the form of 2'-*O*-methyl groups and pseudouridine residues. It is not only the presence but also the location of these modifications that is important. Many of the modifications are situated in what will become functionally important regions of the rRNA. Modifications in the ribosome decoding centre show an increasing reduction in amino acid incorporation as modifications are ablated (Liang, et al., 2009). 2'-*O*-methylation can reinforce or change RNA folding by restricting possible conformations, while pseudouridine modifications can help stabilise RNA structures (Helm, 2006). These modifications are carried out by non-coding RNAs called snoRNAs. The snoRNAs are grouped into the C/D and H/ACA families, forming snoRNPs and utilising the complementarity between themselves and the pre-rRNA to carry out modifications at specific sites. These duplexes are also thought to be required for proper rRNA folding. 43 box C/D snoRNAs are implicated in catalysing 2'-*O*-ribose methylation at 67 sites, while 28 H/ACA snoRNPs catalyse pseudouridylation at 44 sites (Watkins & Bohnsack, 2012).

The Box C/D snoRNP contains Snu13, Nop56, Nop58 with the methyltransferase Nop1 catalysing the nucleotide modification. Box C/D snoRNAs are mainly between 60-90 nucleotides long, characterised by conserved box C (RUGAUGA) near the 5' termini, and box D (CUGA) near the 3' termini (**Figure 2**). These boxes align to form a k-turn motif, containing a noncanonical G-A base pairing, causing a sharp curvature in the RNA structure. This k-turn serves as a binding site for the aforementioned Box C/D snoRNP core proteins, being essential for snoRNA biogenesis and localisation (Kiss-László, et al., 1998). The less well conserved but with the same consensus sequence are the box C' and D' motifs, located more towards the centre of the snoRNA (Kiss, 2001). These conserved structures within the

snoRNA determine the targets for modifications relative to the base paired region. For example in box C/D snoRNAs the target is five bases from the box D or D' motif (Kiss-László, et al., 1996).

The H/ACA snoRNP contains Nhp2, Nop10, Gar1 with the pseudouridine synthase Cbf5 catalysing the nucleotide modification. The H/ACA snoRNAs are between 120 and 140 nucleotides long, containing two hairpins. The hinge in between the two hairpins contains the box H motif (ANANNA), with the box ACA motif being three bases upstream of the 3' end of the snoRNA (**Figure 2**). When directing modifications, the H/ACA snoRNP base pairs to the pre-rRNA with a pseudouridylation loop in the hairpin, with the unpaired uridine in the pre-rRNA being converted to pseudouridine (Ganot, et al., 1997).

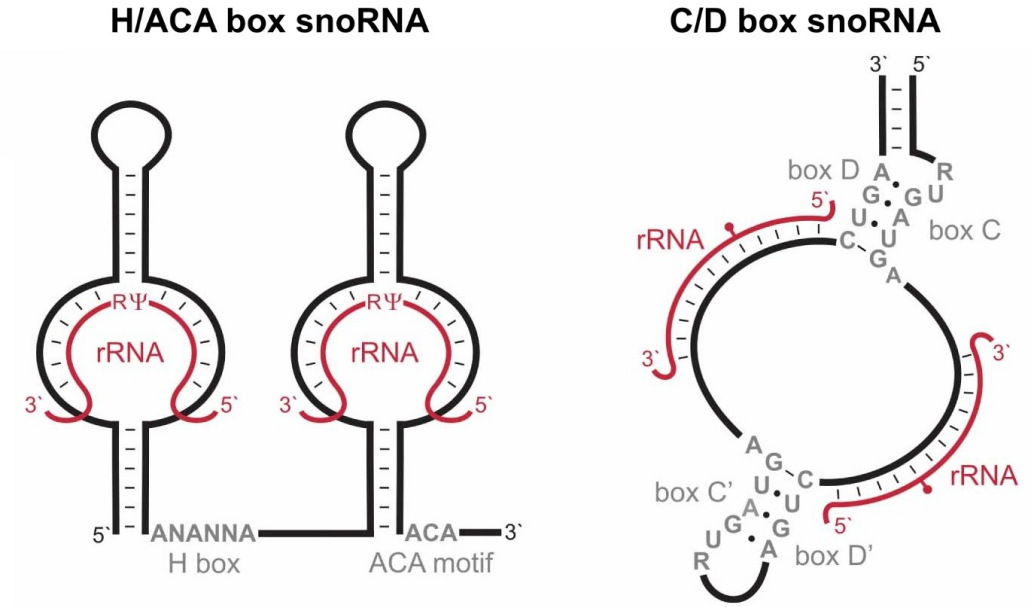


Figure 2, H/ACA and C/D box snoRNAs

Schematic showing the structure of H/ACA and box C/D snoRNAs. The target rRNA sequences are shown in red and conserved box sequences in grey. The pseudouridylated residue for H/ACA box snoRNA is shown by Ψ, and methylated residue for box C/D snoRNA by a red circle. (Adapted from (Watkins & Bohnsack, 2012)).

There are also snoRNPs that are not exclusively involved in modifications, but can also be essential in pre-rRNA folding and processing. These include the box C/D snoRNPs U3, U8 and U14, and the box H/ACA snoRNPs snR30 and snR10.

1.1.2.2 Cleavage of the 35S transcript

The cleavage of the 35S transcript is an essential step in processing. Two internal (ITS) and two external (ETS) transcribed spacer sequences are excised, releasing the 25S, 18S and 5.8S rRNA. These four transcribed spaces are removed by both endonucleolytic and exonucleolytic reactions. These processing steps and progression of precursors are outlined in **Figure 3**. The order of cleavages can change if processed cotranscriptionally, with 70% of nascent transcriptions being cleaved first at A_2 to produce the 20S precursor (Kos & Tollervey, 2010).

If not processed cotranscriptionally, cleavage at A_0, A_1 and A_2 can occur any order, with the resulting intermediates all being detected. Cleavage events at A_0, A_1 and A_2 are dependent on Utp24 (Bleichert, et al., 2006) and Rcl1 (Billy, et al., 2000). Utp24 appears to directly cleave at sites A_1 and A_2 (Wells, et al., 2016), while Rcl1 works in complex with Bms1p (Delprato, et al., 2014) and appears to demonstrate direct endonuclease activity at the A_2 site (Horn, et al., 2011). Cryo-EM maps show that within the SSU these sites are inaccessible to the active site of Utp24, indicating a conformational change occurs to facilitate these cleavage events (Barandun, et al., 2017). Cleavage first at A_0 generates the 33S transcript with subsequent cleavage at A_1 then generating the 32S transcript, which removes the 5' ETS. Cleavage at A_2 then splits the pathway into what will become the mature 18s or 25S/5.8S with the respective 20S and 27SA_2 transcripts (Udem & Warner, 1972). Alternatively cleavage can

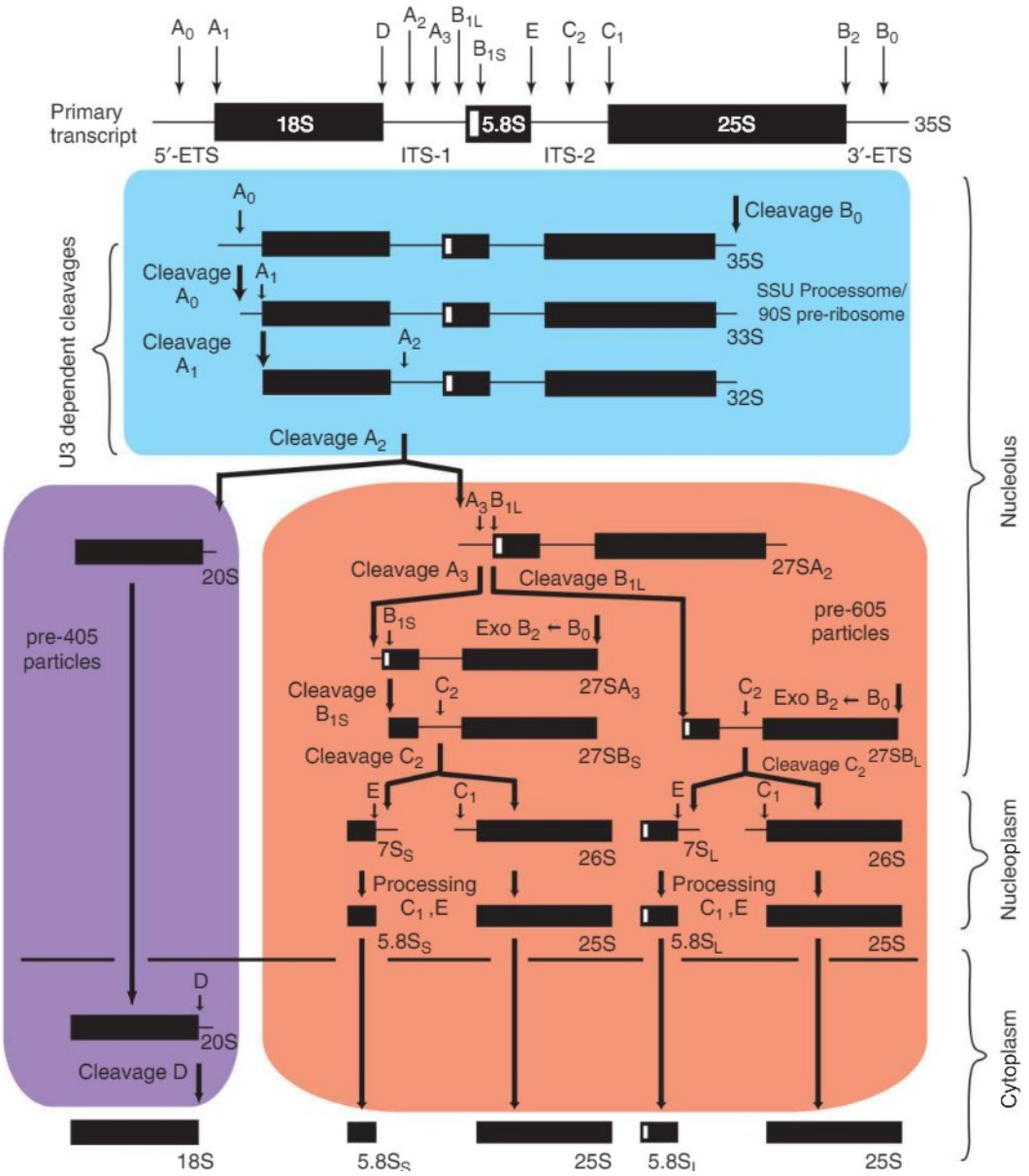


Figure 3, Pre-rRNA processing events

Schematic illustrating the different stages of pre-rRNA processing. Highlighted in blue are the U3 dependant cleavage events (A_0, A_1, A_2), in purple the pre-40S stages and in orange the pre-60S stages. The different cleavage sites are indicated, along with their location in the cell, detailing the entire progression of cleavage steps required to obtain the mature rRNA. (From (Phipps, et al., 2010))

occur at another site in the ITS1, the A_3 site, to produce the 23S and $27SA_3$ transcripts (Lygerou, et al., 1996). It was found that yeast may switch to the A_3 pathway in response environmental stresses such as nutrient availability (Kos-Braun, et al., 2017). This pathway is non-productive, stalling the small subunit progression and halting ribosome production. The U3 snoRNA is essential to these A_0, A_1 and A_2 cleavage events, thought to help in the folding and timing during these events, binding in the 18S and 5' ETS of the pre-rRNA.

After cleavage at site A_2, around 85-90% is then cleaved at A_3 to produce the $27SA_3$. The residual ITS1 is then removed by 5'-3' exonuclease Rat1 and Rrp17, stopping at site $B1_S$ to create the 5' end of 27SBs. The 10-15% that don't follow this pathway are cleaved directly at $B1_L$, resulting in the $27SB_L$ intermediate. Both $27SB_S$ and $27SB_L$ are processed identically from this point, packaged into the pre-60S particle, both resulting in the mature 25S, and $5.8S_S$ or $5.8S_L$ respectively. The 20S is the other intermediate created by the A_2 cleavage, which is then packaged into the pre-40S particle.

1.1.1 The U3 small nucleolar RNA

U3 snoRNA was the first snoRNA identified (Weinberg & Penman, 1968) , and is perhaps the best studied. Unlike the other 60-90 nucleotide C/D box snoRNAs U3 snoRNA is 333 nucleotides in length and does not direct chemical modifications. Due to its essential role in ribosome biogenesis it is one of the most abundant RNA molecules, with 400-1000 copies per yeast cell. It is a critical trans-acting RNA-chaperone, with knockout studies implicating with the essential A_0, A_1 and A_2 cleavage events (Kass, et al., 1990) as the corresponding pre-rRNA intermediates accumulate (Hughes & Ares, 1991). It is also structurally important to the SSU processome, dictating progression via complementarity with the 5' ETS and 18S regions of the pre-rRNA (Dutca, et

al., 2011 , Barandun, et al., 2017). U3 snoRNA archives these roles while part of the U3 snoRNP with Nop1, Snu13, Nop56, Nop58, Rrp9, and also associates with the Mpp10 complex (Imp3, Imp4, Mpp10). This unique C/D box RNA is thought to be ubiquitously present in eukaryotes, and despite high sequence variation it displays structural conservation with phylogenetically conserved box A', A, C', B, C and D motifs (Marz & Stadler, 2009).

U3 snoRNA is present in two copies in the yeast genome, snR17a on chromosome XV and snR17b on chromosome XVI. The two genes share 98% similarity, but snR17a 5-10 is times more abundant. A single knockout of either gene is tolerated, requiring a double knockout for lethality. Both copies contain an intron, with an atypical branch point (GACTAAC) (Myslinski, et al., 1990). It is expressed from a promoter with two elements previously thought to be mutually exclusive, the Homol D-box and TATA-box (Nabavi & Nazar, 2008). This promoter peculiarity reflects U3 snoRNAs unique role compared to other snoRNAs. The mature U3 snoRNA has a 5' cap structure which is hypermethylated to a 2,2,7-trimethylguanosine (TMG) cap, which is reliant on the box C/D motif (Speckmann, et al., 2000).

The biogenesis of U3 snoRNP occurs in the nucleoplasm and involves a large multiprotein complex, with assembly and processing occurring concurrently (Watkins, et al., 2004). The assembled snoRNP complex is then transported to the nucleolus (Verheggen, et al., 2002). The U3 snoRNP is found in both the DFC and GC (nucleolus compartments), consistent with its role in ribosome biogenesis in these regions. The transition of U3 snoRNP to the GC is concomitant on SSU processome progression, with defects in SSU processome formation or ablating Mpp10 binding leading to U3 accumulating in the DFC (Granneman S, 2004). It was found that the C'/D box motifs but not

B/C is essential for nucleolar localisation, but either are sufficient for nuclear retention (Narayanan, et al., 1999 , Speckmann, et al., 1999).

The structure of U3 snoRNA was initially resolved by chemical probing studies (Méreau, et al., 1997) and mutagenesis studies (Samarsky & Fournier, 1998). It can be divided into three domains: the 5' domain, hinge regions and the 3' domain (**Figure 4A**). The 5' domain contains the GAC box, box A' and box A motifs, containing complementary sequences to the pre-18S buried in a stem loop that must be unwound (Gérczei & Correll, 2004). The hinge region contains the unstructured 3' hinge and 5' hinge with a stem-loop in-between, and contains complementary sequences to the 5' ETS region of the pre-rRNA. The secondary structure of the U3 snoRNA when binding the pre-rRNA shown in **Figure 4B**. The 3' region contains the box C', box D, box B and box C motifs, where the U3 snoRNP proteins bind. These protein interactions consequently make these regions important to pre-rRNA processing, RNA stability, cap hypermethylation, nucleolar retention and localisation (Speckmann, et al., 2000 , Lafontaine & Tollervey, 1999). Mutations in the C'/D but not B/C destabilise U3 snoRNA, with the C'/D box being essential for accumulation (Samarsky & Fournier, 1998). There are many distinct differences observed when disrupting the C'/D and B/C box motifs, reflecting the alternative assortment of proteins that associate with either one. The 3' domain stem III separates the box motifs, with three protruding stems (stem IV, V, VI), all with multiple internal loops and bulges. Despite making up a large proportion of the molecule, the majority of yeast/fungal specific stem IV and VI are not functionally or structurally essential (Samarsky & Fournier, 1998). The box C'/D and box B/C motifs are on opposite sides of a stem loop, forming k-turn motifs important for protein interactions. K-turn motifs consist of a 3 nucleotide bulge, flanked by essential helices (Klein, et al., 2001).

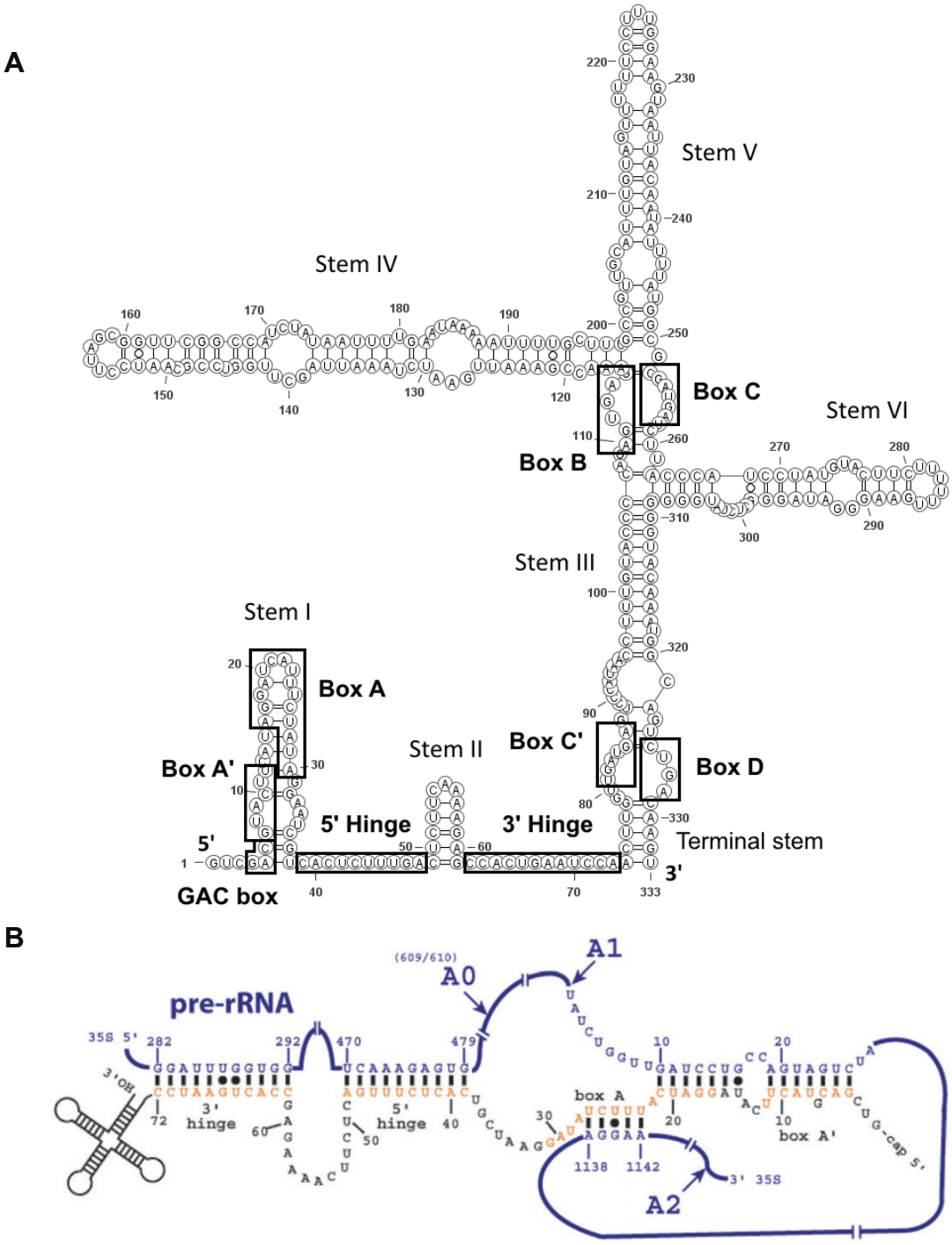


Figure 4, U3 small nucleolar RNA

A. Schematic showing the secondary RNA structure of U3 snoRNA, with the bases and base pairing indicated. The conserved box motifs are highlighted (GAC box, Box A', Box A, Box C', Box D, Box B, Box C), along with the hinge regions (3' hinge and 5' hinge).
B. Schematic of the 5' end of U3 snoRNA with conserved regions (orange) forming duplexes with the 35S pre-rRNA (blue), with the cleavage events indicated, participating in folding the pre-18S transcript into the central pseudoknot. (Adapted from (Henras, et al., 2008)).

Chemical probing indicates the initial base pairing occurs between nucleotides 62-71 in the 3' hinge of U3 snoRNA and nucleotides 281-291 in the 5' ETS. (Dutca, et al., 2011). As expected this interaction is essential for growth, and a prerequisite for the subsequent addition of factors (Sharma & Tollervey, 1999 , Marmier-Gourrier, et al., 2011). The initial 3' hinge hybridisation is essential for subsequent 5' hinge hybridisation to nucleotides 447-479 of the 5' ETS, but not vice versa, also indicating the 3' hinge is the preliminary interaction (Dutca, et al., 2011). No protein interactions are essential for the initial 3' hinge interaction *in vitro* (Shah, et al., 2013). The 5' hinge interaction is essential for pre-rRNA processing and thus growth (Beltrame & Tollervey, 1995). Nucleotides 16-22 within the box A' hybridise with nucleotides 9-15 of the 18S pre-rRNA, and nucleotides 3-13 of GAC box/box A' hybridise with nucleotides 1111-1122 of the 18S pre-rRNA. The secondary structure formed by this complementarity is incompatible with the formation of the 18S into a central pseudoknot, a conserved RNA structure critical to the coordination of the 40S subunit (**Figure 4B**). In this way U3 snoRNA can stall progression. An interaction between nucleotides 304-315 of U3 snoRNA and two regions of the 18S pre-rRNA (1,063-1,073 and 1,624-1,643) was detected by cross-linking, ligation, and sequencing of hybrids (CLASH) (Kudla, et al., 2011). This is predicted to be close to the mature central pseudoknot, and is situated where other box C/D snoRNAs have their modification guides sequences. This however would involve disrupting the essential stem III (Samarsky & Fournier, 1998), making unclear how this interaction is feasible.

The duplex formation and sequestration of the conserved sequences in the 18S allows U3 snoRNA to stall progression and prevent premature pseudoknot maturation. The cleavage sites A_0 and A_1 are also neighbouring this area of complementarity, and are reliant on the presence of U3 snoRNA. The 3' and 5' hinges of U3 snoRNA hybridise to the 5' ETS region of the pre-rRNA to provide rigidity to the SSU processome structure. The hinge-5' ETS base

pairing is situated in the core of the SSU processome particle, and is utilised as a structural scaffold for complexes such as UtpA and UtpB (Chaker-Margot, et al., 2017). The hinge base pairing also dictates constraints upon the pre-rRNA, which influences the 18S regions that bind to the box A and A' motifs. By base pairing with the pre-rRNA, U3 snoRNA is able to manipulate the topology of the SSU processome through restriction of conformational changes (Dutca, et al., 2011 , Barandun, et al., 2017).

1.1.1.1 U3 ribonucleoproteins and the Mpp10 complex

U3 snoRNA accomplishes its essential role both via proteins that interact with the B/C and C'/D motifs to form the U3 snoRNP, and with proteins within the SSU processome (Dragon, et al., 2002 , Barandun, et al., 2017). The SSU processome that forms at the 5' terminal knob of the nascent 35S pre-rRNA transcript consists of ~100 factors, many of which directly or indirectly interact with the U3 snoRNP (Vincent, et al., 2018). U3 snoRNA appears to be a central organising player in the SSU processome, situated at the core and reaching to the outside (Barandun, et al., 2017). The U3 snoRNP consists of the same RNP proteins as other box C/D snoRNAs: Nop1, Nop56, Nop58, Snu13 and U3 snoRNP specific Rrp9 (Watkins, et al., 2000 , Venema, et al., 2000). In accordance with other box C/D snoRNAs there are two copies of Nop1 and Snu13. Rrp9 is dependent on the initial binding of Snu13 to the k-turn of the box B/C motif (Granneman, et al., 2002), preceding Nop1 and Nop56. The other copies of Snu13 and Nop1 along with Nop58 bind the C'/D motif, with a Nop56-Nop58 heterodimer connecting the two k-turn motifs (Cahill, et al., 2002) (Chaker-Margot, et al., 2017 , Barandun, et al., 2017). Cryo-EM has allowed many assumptions about the organisation of both the U3 snoRNP and surrounding SSU processome to be validated.

Nop56 and Nop58 together form a pseudosymmetric heterodimer with their helical domains, creating a binding surface to grip the U3 snoRNA. The N-terminal domain of Nop56 and C-terminal domain of Nop58 each interact with one pair of Nop1 and Snu13 (Gautier, et al., 1997) (Cheng, et al., 2017) In this way Nop56 and Nop58 link the box B/C and box C'/D motifs. Nop56 also stabilises the interaction between the 5' hinge and 5' ETS, while Nop58 stabilises the 3' hinge interaction (Cheng, et al., 2017). Nop58 is implicated in both box C/D snoRNA biogenesis and stability. This is shown by depletion both under an inducible promoter (Lafontaine & Tollervey, 1999) and siRNA knockdown (Watkins, et al., 2004) leading to dramatic loss of box C/D snoRNAs including U3 snoRNA. Nop56 knockouts affect pre-rRNA processing, but are not as important as Nop58 for box C/D snoRNA accumulation (Lafontaine & Tollervey, 2000). Although Nop56 and Nop58 are core proteins of all box C/D snoRNAs, modifications are much less affected than pre-rRNA processing when depleted (Lafontaine & Tollervey, 1999). This may reflect modifications only necessitating a transient interaction relative to that required for pre-rRNA processing.

Rrp9 is a protein uniquely associated with U3 snoRNA, essential for pre-rRNA cleavage events. Depletion of Rrp9 causes U3 snoRNA specific defects, while other snoRNAs remain unaffected. The box C motif is essential to binding, and the box B motif has supportive role (Venema, et al., 2000). Rrp9 contains WD40 repeats, with a seven-bladed propeller fold. A 7bc loop atop the propeller is important for Rrp9 to recognise the B/C motif (Zhang L, 2013). A study with human U3 snoRNA found that hU3-55K (Rrp9) required the initial binding of 15.5K (Snu13) in order to associate with the box B/C motifs, highlighting the stepwise assembly (Granneman, et al., 2002). 15.5K was found to bind the box B/C motifs with higher affinity than the box C'/D motifs, and is essential to proper pre-rRNA processing (Watkins, et al., 2000). Mutating flanking nucleotides indicated 15.5K was dependent on the B/C

motifs, but it was the surrounding structure, not sequence, that were important to the recognition by hU3-55K (Granneman, et al., 2002). Using systematic evolution of ligands by exponential enrichment (SELEX) the essential elements were identified for Snu13 and Rrp9 binding. (Cléry, et al., 2007). The ability to form a k-turn structure was essential for Snu13 binding, as were the flanking helices, as previously shown for box C/D snoRNAs (Szewczak, et al., 2005). Also interestingly, changes that increased Snu13 affinity, abolished Rrp9 binding, highlighting that conformational changes subsequent to Snu13 binding are important for Rrp9 binding (Granneman, et al., 2002). An internal loop was important in human U3 snoRNA to hU3-55K binding, however these loops have been shown to not be essential in yeast, indicating the binding of Rrp9 may not be identical to hU3-55K. A major structural difference between human and yeast U3 snoRNA is a large stem flanking to the box B/C motifs, that Rrp9 has been crosslinked with (Granneman, et al., 2009). This stem may be contribute to the differences in binding site for Rrp9 in yeast. This is also reflected in the protein differences between Rrp9 and hU3-55K, with hU3-55k only partially complementing in yeast (Venema, et al., 2000). Another difference with human U3 snoRNA is a 'weak' C' box, that leaves assembly reliant on hU3-55K binding. (Knox, et al., 2011). Yeast does not contain the 'weak' C' box and thus accumulation is not effected by Rrp9 depletion.

Mpp10 was one of the first yeast U3 snoRNP proteins identified (Dunbar, et al., 1997), and has been used as a bait to characterise further factors (Dragon, et al., 2002). Unlike other proteins that were found to associate to the 3' domain of U3 snoRNA, Mpp10 was found to be dependent on sequences in the hinge region (Wormsley, et al., 2001), implicating it's role in pre-rRNA duplex formation. A cold sensitive truncated mutant of Mpp10 implicated it in cleavage events A_1 and A_2, but not A_0 (Lee & Baserga, 1997). The Mpp10 complex binds after transcription of the A_0 site to complete the 5'-ETS particle. Cryo-EM maps have highlighted multiple roles within the SSU processome, containing a

flexible N-terminus and structured C-terminus, involved in stabilisation and bridging. Mpp10 unfolds and stabilises helix 44 in the pre-18S, keeping it isolated until it is ready to interact with rRNA domains that form the central pseudoknot.

Imp3 and Imp4 are also essential to pre-rRNA processing but not U3 snoRNA levels, and found to be associated with Mpp10 (Lee & Baserga, 1999), and together bind to form the Mpp10 complex (Granneman, et al., 2003). This complex is situated at the core of the SSU processome near the 5' domain of the U3 snoRNA, with Imp3 and Imp4 providing a binding platform for Mpp10 which then completes the 5'-ETS particle. The N terminus of Imp3 stabilises both the 5' hinge interaction with 5' ETS, and the proceeding nucleotide linker connecting to the box A. Imp3 and Imp4 are essential to the association and dissociation of both the U3-ETS and U3-18S duplex (Gérczei & Correll, 2004). While the initial 3' hinge interaction requires no proteins *in vitro*, the 5' hinge interaction requires both imp3 and imp4 to satiate *in vivo* demand. The U3-18S interaction involves both Imp3 and Imp4, but only Imp3 is essential due to its increased efficiency in unfolding the stem loop structure around the box A and A' motif. This unfolding is required as the complementary bases are buried in the stem loop. Imp3 also unfolds bases in the 18S pre-rRNA region, promoting hybridisation (Shah, et al., 2013). Imp3 is vital in stabilising the unfolded forms, as refolding is more favourable than forming the U3-18S duplex. *In vitro* studies showed that addition of Imp3 and Imp4 increased U3-18S duplex formation from an undetectable rate of ≤400 M^{-1} s^{-1} to a rate of $\sim 10^6 M^{-1}$ s^{-1}, which is the same as the spontaneous rate between short unfolded duplexes (Gérczei, et al., 2009). Another role of Imp4 is destabilising the U3-18S duplex when the SSU is ready to progress, facilitating the rapid turnover needed in ribosome biogenesis. *In vitro* dissociation rates are 16-fold higher in the presence of Imp4. It's action is insufficient to support *in vivo* demand however, indicating it may complement

to the role of other helicases such as Dhr1 or Has1. Dhr1 was shown to unfold the duplex at box A but not A' (Sardana, et al., 2015), so Imp4 may contribute to box A' unfolding. Dhr1 is partially required to catalyse the unwinding of the U3-18S duplex in order to maintain rapid rates of ribosome assembly (Sardana, et al., 2015). Without the action of Dhr1 the spontaneous dissociation rate is slow to satiate *in vivo* demand. Before the A_1 cleavage event, cryo-EM shows the site to be inaccessible to Dhr1, requiring conformational change (Barandun, et al., 2017).

Nop1 is a methyltransferase found in two copies in the U3 snoRNP, essential for pre-rRNA processing (Tollervey, et al., 1991) and cell viability (Schimmang, et al., 1989). It is also essential for the stability and accumulation of certain snoRNAs (not U3 snoRNA) (Lafontaine & Tollervey, 2000). Unlike in other box C/D snoRNPs, it has not been implicated in 2'-O-methylation role as part of U3 snoRNA, and has seemingly evolved instead to serve a structural scaffolding role. This may indicate that U3 snoRNP is an example of a particle whose original enzymatic activity has been usurped by an architectural role. This is reflected by Nop1 assuming an alternative conformation in U3 snoRNP, different than found in other box C/D snoRNPs (Lin, et al., 2011 , Kornprobst, et al., 2016). Using UV crosslinking Nop1 was found to bind near a region of U3 snoRNA (positions 299-302, 315-317) that was also found to bind the 18S pre-rRNA (positions 304-315) (Kudla, et al., 2011). This region is near the modification guide sequence of other box C/D snoRNAs. This could potentially represent U3 snoRNA's original modification guide sequence. Within the SSU processome one copy of Nop1 serves as a binding platform for Fcf2, Sas10 and Utp24 and the other serves as a binding platform for Utp11 and Bud21.

1.1.2 Ribosome misassembly

Despite the essential role of ribosomes to cell survival, there are viable mutations that cause ribosomopathies. Dysfunction in ribosome biogenesis has been commonly linked to human disease, with rates being upregulated in most cancers due to its direct link to proliferative potential. (Ruggero & Pandolfi, 2003). Mutations in the SSU processome stage of ribosome biogenesis have been linked to multiple ribosomopathies (Sondalle & Baserga, 2014). Missense mutations in UTP4 are implicated in North American Indian childhood cirrhosis, truncations of UTP14 are linked to infertility and a missense mutation in EMG1 causes Bowen-Conradio syndrome. The cell type specific defects on a ubiquitous process like protein synthesis may indicate cell specific ribosome components, or differing ribosome defect detection mechanisms. Other mutations in ribosome biogenesis factors have been linked to inherited bone marrow failure syndromes, seemingly due to precursor red bloods cells having a large demand for ribosomes (Ganapathi & Shimamura, 2008). Many of these conditions also have an increased risk for cancer development as well as the specific symptoms. The snoRNAs U22, U8, U94, U97 and U3 were all found in increased levels in both breast cancer cells lines and breast cancer primary cultures (Su, et al., 2014). Mutations in the U3 interacting proteins have also been linked to various carcinomas (Sjöblom, et al., 2006).

1.2 Genetic interactions and fitness landscapes

A fitness landscape is an elegant way to conceptualise evolutionary trajectories, depicting the relationship between genotype and phenotype, with its topology consisting of high fitness peaks and low fitness valleys (Wright, 1932). The specific dimensions of a particular landscape are determined by selective pressures, with multiple peaks representing alternative evolutionary potentials. Selection will drive a population to the local optima, with the acquisition of mutations facilitating the climbing of the peak. Populations may be unable to reach the global fitness maxima if it requires traversing a low fitness valley. Two evolutionary trajectories are illustrated in a hypothetical fitness landscape in **Figure 5**.

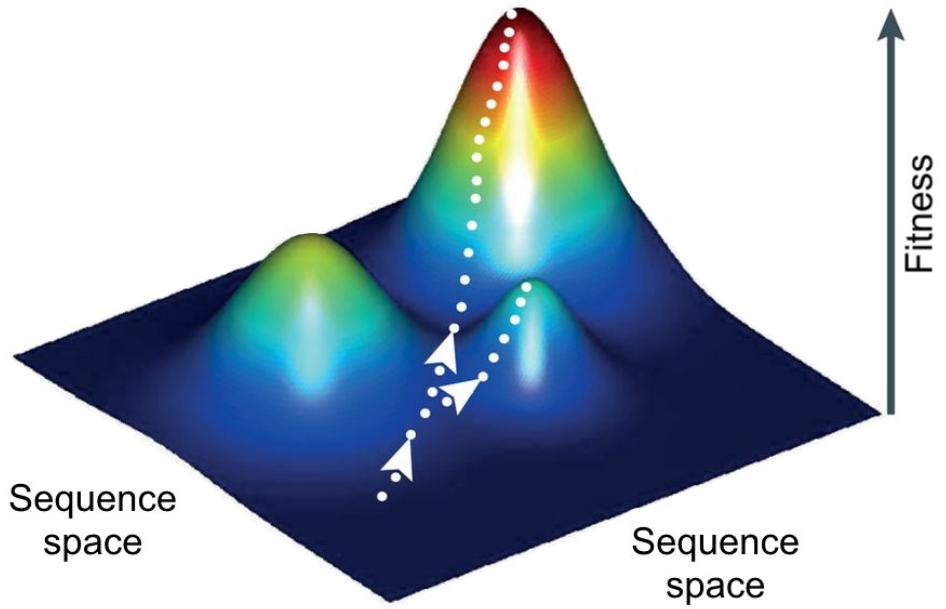


Figure 5, Rugged fitness landscape

A 3D visualisation of a fitness landscape, with three peaks divided by low fitness valleys. Two hypothetical evolution trajectories are shown with the white dotted lines, each committing to a different peak. The XY plane represents sequence space, and the Y plane represent fitness. (Adapted from (de Visser & Krug, 2014)).

The molecular basis behind traversing a fitness landscape involves the accumulation of mutations effecting function, with different trajectories being viable. It seems most molecules are well adapted, with advantageous mutations being rare (Eyre-Walker & Keightley, 2007). When looking at the trajectories of bacterial β-lactamase adaptation to an antibiotic, the accumulation of negative or intermediate neutral mutations was common en route to optimal resistance (Weinreich, et al., 2006 , Poelwijk, et al., 2007). The combination of different paths, some containing only beneficial mutations and some with conditionally beneficial mutations, results in a complex topography. Detrimental and neutral mutations within a path to resistance make it less accessible to evolution. An example found involved a mutation which increased enzyme activity at the cost of stability, and a mutation which increased stability (Weinreich, et al., 2006). Initially acquiring the mutation that lead to increased stability was neutral, but could compensate for the reduction in stability when increasing enzyme activity.

1.2.1 Epistasis

Epistasis (genetic interaction) describes the interactions between mutations, where the phenotype differs from the additive expectation (Phillips, 1998). This encompasses circumstances where the genetic background in which a mutation occurs is influential. In its simplest form this involves the interaction between pairs of mutations. Two mutations may both inactivate the same pathway, presenting a more modest sum effect than expected (coequal interaction). Conversely, two mutations independently may have a subtle effect on a pathway due to redundancy, whereas together they may substantially disrupt it (negative interaction). Different scenarios are outlined in **Figure 6**, illustrating the consequence of the sequential accumulation of two mutations. These represent just four potential scenarios, each having many potential biological explanations. When no epistasis exists the effect of genetic

background has no effect, with the two mutations being independent of one another. For magnitude epistasis the magnitude of the effect is increased by the presence of either mutation, with the two mutations potentially compromising the robustness of a certain pathway. This is termed sign epistasis when mutations are only conditionally beneficial, requiring both mutations to improve fitness. Reciprocal sign epistasis represents when both mutations cause a change in the sign of the fitness effect, being individually detrimental but beneficial when combined. The fitness cost of reciprocal sign epistasis is a major source of fitness landscape ruggedness, leading to populations being trapped in suboptimal peaks due to the requirement of both mutations.

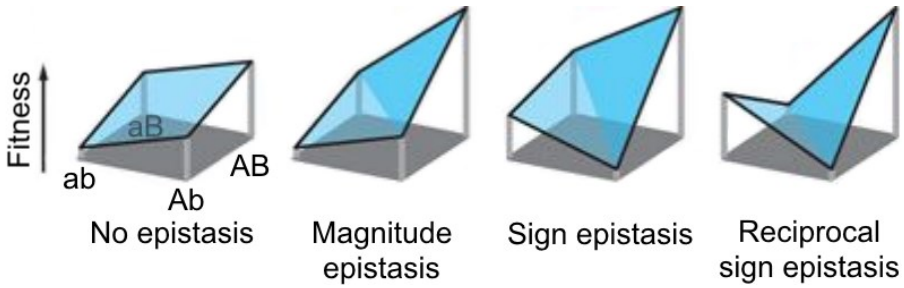


Figure 6, Instances of epistasis

Four examples of epistasis, from genotype ab to genotype AB. For each example to two possible indeterminate steps are shown (aB or Ab), with the fitness implications on the y axis (Adapted from (Poelwijk, et al., 2007)).

Pairwise epistasis does not occur in a vacuum, and thus the genetic background in which they exist can have a dramatic influence. This interaction of multiple mutations is named higher-order epistasis, and contributes to change the topology to a rugged multi-peak fitness landscape. This is termed intramolecular with interactions between one gene, and intermolecular between multiple genes. An example of intergenic epistasis could involve

mutations in two protein binding partners, with it being an example of reciprocal epistasis if both mutations are required to avoid negative fitness effects and increase fitness. These higher order interactions, both intramolecular and intermolecular, are thought to be widespread in nature in functionally related networks. When examining the effect of higher order epistasis on 16 empirical fitness landscapes, it was found that the effect upon a landscape increases with increasing order (Weinreich, et al., 2018).

1.2.1 Genetic and environmental interactions

Mapping genotype to phenotype involves nonlinear interactions, involving both environmental and genetic factors. This contributes significantly to the complexity of studying disease susceptibility genes (Moore, 2003). Polymorphisms may be identified that are only significant in a specific genetic background. The influence of the background in which mutations occur has been recognised for over 100 years, however is still hard to quantify. The presence of these interactions is thought to be prevalent, as any eukaryotic gene can be regulated by 100s of proteins. This is also illustrated in the omnigenic model of complex traits, whereby seemingly the whole genome can have an influence upon a trait (Boyle, et al., 2017).

The environment in which mutations occur can dictate the optimality of peaks, with changing environments transitioning populations to areas of higher or lower fitness (Flynn, et al., 2013). The changing of environments can also provide temporary bridges to span low fitness valleys (Steinberg & Ostermeier, 2016). The study and quantification of gene-gene (intergenic epistasis) and gene-environment interactions is a complicated but key step in understanding the genetic architecture of gene expression and human disease (Wu & Ma, 2018).

The study of gene-gene interactions is a topic currently widely being pursued both with more artificial means through high-throughput mutational studies (Costanzo, et al., 2010), and in the study of human traits (Hemani, et al., 2014). Gene-environment interactions can also be detected in various human diseases such as cancers (Hunter, 2005) and also at a molecular level, where for example the fitness impact of mutations within a riboenzyme were dependent on the environment (Hayden & Wagner, 2012). The influence of gene-environment interactions on biological processes in the context of diseases is an area of great interest to better understand and treat them (Hunter, 2005), with it being difficult to dissect apart the influence of genotype and environmental factors (Ottman, 1996).

1.2.2 Empirical fitness landscapes

Studying the genotype-phenotype relationship experimentally with empirical fitness landscapes can systematically plot evolutionary trajectories (de Visser & Krug, 2014). Predicting the available paths to antibiotic resistance can provide a clearer explanation on how it is achieved. When looking at trimethoprim resistance it was found there was a multi-peaked adaptive fitness landscape, with epistasis having a large influence on delaying and obscuring which peak a population is committed to (Palmer, et al., 2015).

Due to the recent advancements in genomic technologies, the characterisation of an entire gene's fitness landscape is possible. This provides the fitness impact of mutational variants within a gene. To systematically create a complete fitness landscape of a gene requires testing a broad range of mutational phenotypes. This can be achieved by deep mutational scanning, which combines mutagenesis, gene replacement and next generation sequencing. By enriching for a particular phenotype, thousands of mutant

variants may be assayed. The high-throughput nature provides a tremendous advantage over the traditional procedure of characterising individual variants. This method can be applied to both proteins (Araya & Fowler, 2011) and RNA (Pitt & Ferré-D'Amaré, 2010), providing data at a single amino acid/nucleotide resolution. In the case of proteins the epistatic interactions have been used to accurately predict protein structure (Schmiedel & Lehner, 2018 , Rollins, et al., 2018). The limitations come in experimental design, constrained by the feasibility of screening for a particular phenotype. There are also considerations to be made when designing the parameters of the experiments to ensure the fidelity of the fitness estimates (Matuszewski, et al., 2016).

1.3 Aims of this thesis

The interactions both within a genome and with the surrounding environment play a significant role in shaping phenotypic variation. Investigating these interactions in the context of a whole molecule's fitness landscape is largely unexplored. Characterising these interactions will give insights into the underlying molecular biology. In this thesis I systematically and quantitatively survey these interactions within the fitness landscape U3 snoRNA, using a high-throughput fitness assay.

In this thesis I aim to tackle the following objectives:

- **Survey gene-gene interactions within the fitness landscape of U3 snoRNA with U3 snoRNA interacting proteins.**

- **Survey gene-environment interactions within the fitness landscape of U3 snoRNA in different temperature environments.**

2 MATERIALS AND METHODS

2.1 Sources of reagents

Name of Reagent	Company	Use	Cat. No.
QIAprep Spin Miniprep Kit	Qiagen	Plasmid extraction	27104
MinElute PCR Purification Kit	Qiagen	DNA purification	28004
QIAquick PCR Purification Kit	Qiagen	DNA purification	28104
QIAGEN Plasmid Plus Maxi Kit	Qiagen	Plasmid extraction	12963
TURBO DNA-free Kit	Thermofisher	DNAse treatment	AM1907
Deoxyribonucleic acid sodium salt from salmon testes	Sigma-Aldrich	Yeast transformation	D1626-250MG
Acid-Phenol:Chloroform, pH 4.5 (with IAA, 125:24:1)	Life Technologies	RNA extraction	AM9720
Phenol:Chloroform, pH 6.6 (25:24:1)	Life Technologies	DNA extraction	AM9730
1-Naphthaleneacetic acid	Sigma-Aldrich	Use in auxin inducible degron system	N0640-25G
Chloroform	Sigma-Aldrich	RNA/DNA extraction	15771819
Phusion High-Fidelity PCR Master Mix with HF Buffer	New England Biolabs	PCR	M0531L
AccuPrime Pfx SuperMix	Thermofisher	PCR	12344040
NuPAGE MOPS SDS Running buffer	Thermofisher	Western blotting	NP0001
NuPAGE 10% Bis-Tris Protein Gels, 1.0 mm, 12-well	Thermofisher	Western blotting	NP0302BOX
NuPAGE LDS Sample Buffer (4X)	Thermofisher	Western blotting	NP0007
SuperScript III Reverse Transcriptase	Thermofisher	Generate cDNA	18080044
LightCycler® 480 SYBR Green I Master	Roche	qPCR	04707516001
Restriction enzymes (EcoR1, Sal1, Dpn1, BamH1, Kpn1, HindIII, Not1, Sac1, EcoRV, Not1)	New England Biolabs	Restriction digests	

Oligonucleotides	IDT	Cloning, sequencing	
Bacto-peptone	Becton Dickinson	Yeast media	211677
Yeast extract	BD Difco	Yeast media	212720
Yeast nitrogen base without amino acids	Sigma-Aldrich	Yeast media	Y0626-1KG
Complete synthetic medium without adenine	MP Biomedicals	Yeast media	DSCK1001
Galactose	Sigma-Aldrich	Yeast media	G5388-1KG
Glucose	Sigma-Aldrich	Yeast media	G7021-1KG
Agar	Sigma-Aldrich	Yeast media	79883-500G
E-Gel EX Gel, 2%, 20-Pak	Thermofisher	Gel extraction	G402002
Q5 polymerase	New England Biolabs	PCR	M0491
iBlot 2 Transfer Stacks, nitrocellulose, mini	Life Technologies	Western blotting	IB23002
Pierce Supersignal Femto kit	Thermofisher	Western blotting	34095
PerfectHyb Plus Hybridization buffer	Sigma-Aldrich	Northern blotting	H7033-1L

2.2 Antibodies

Antigen	Lot num	Organism	Company	Use	Dilution	Cat. No
FLAG	SLBT6752	Mouse	Sigma	Western	1:2000	F3165-.2MG
alpha Tubulin	GR156722-6	Rabbit	Abcam	Western	1:2000	ab184970
Anti-rabbit		Goat	Abcam	Western	1:10000	ab205718
Anti-mouse		Goat	Abcam	Western	1:10000	ab6728
Anti-mini-AID-tag	003	Mouse	MBL	Western	1:1000	M214-3

2.3 Growth media and buffers

Yeast and bacterial media was autoclaved by the MRC HGU IGMM technical services and then stored at room temperature for one month.

*Prepared by technical services at MRC HGU IGMM

Name	Component	Concentration
Yeast media		
YPGalactose +ADE media	Yeast extract	1%
	Bacto-peptone	2%
	Galactose	2%
	Adenine sulfate	0.003%
YPGalactose +ADE x2 media	Yeast extract	2%
	Bacto-peptone	4%
	Galactose	4%
	Adenine sulfate	0.003%
YPGalactose plates	Yeast extract	1%
	Bacto-peptone	2%
	Galactose	2%
	Agar	2%
YPGalactose G418 plates	Yeast extract	1%
	Bacto-peptone	2%
	Galactose	2%
	Agar	2%
	G418	200 µg/ml
YPGalactose Hygromycin plates	Yeast extract	1%
	Bacto-peptone	2%
	Galactose	2%
	Agar	2%
	Hygromycin B	200 µg/ml
CSM -ADE Galactose media	YNB without amino acids	According to instructions
	CSM without adenine	According to instructions
	Galactose	2%
CSM -ADE Galactose plates	YNB without amino acids	According to instructions
	CSM without adenine	According to instructions
	Galactose	2%
	Agar	2%

CSM -ADE Galactose + NAA media	YNB without amino acids	According to instructions
	CSM without adenine	According to instructions
	Galactose	2%
	NAA	100-1000 µM
CSM -ADE Galactose + NAA plates	YNB without amino acids	According to instructions
	CSM without adenine	According to instructions
	Galactose	2%
	NAA	100-1000 µM
	Agar	2%
CSM -ADE Glucose media	YNB without amino acids	According to instructions
	CSM without adenine	According to instructions
	Glucose	2%
CSM -ADE Glucose + NAA media	YNB without amino acids	According to instructions
	CSM without adenine	According to instructions
	Glucose	2%
	NAA	100-1000 µM
1% Potassium acetate + ¼ required ADE	Potassium acetate	1%
	Adenine sulfate	0.00075%
Bacterial media		
* LB	Bacto-tryptone	1%
	Yeast extract	0.5%
	NaCl	0.5%
* SOC	Bacto-tryptone	2%
	Yeast extract	0.5%
	NaCl	0.06%
	KCl	0.02%
	MgCl2	0.1%
	MgSO4	0.12%
	Glucose	0.4%
* LB ampicillin plates	Bacto-tryptone	1%
	Yeast extract	0.5%
	NaCl	0.5%
	Agar	1.5%
	Ampicillin	50µg/ml
Buffers		
* PBS	NACl	140 mM
	KCl	3 mM
	Na2HPO4	8 mM
	H2PO4	1.4 mM
	MgCl2	20 mM

* TE	Tris base	2 M
	Acetic acid	5.71%
AE buffer	Sodium acetate	50 mM
	EDTA	10 mM
RNA loading dye	Ficoll	1.5g
	Bronophenol blue	25mg
	Xylenocyanol	25mg
Northern blot stainer (100mL)	0.3M NaAc pH 5.3	10M
	Methylene blue	20mg
NBC (x10) (10L)	Boric acid	30.92g
	NaAc	0.82g
	NaOH	2g
20x SSC (1L)	NaCl	175.3g
	Sodium citrate dihydrate	88.2g
KPi (potassium phosphate) buffer A, pH 6.5	KH_2PO_4	1.9 g/L
	K_2HPO_4	1.05 g/L
KPi (potassium phosphate) buffer B (buffer A + zymolyase), pH 6.5	KH_2PO_4	1.9 g/L
	K_2HPO_4	1.05 g/L
	Zymolyase	2 g/L
PEG 50% (100ml)	PEG	50g
* 20x TBE	Tris Base	216 g/l
	Boric Acid	110 g/l
	0.5M EDTA, pH 8	80 ml/l
10x TBS-T	Tris base	12.114g
	NaCl	43.83
	Tween 20	2.5ml
	H_20	To 500ml

2.4 Oligonucleotides

Ordered from IDT. All standard purification, apart from for next-gen sequencing related oligonucleotides then HLPC was used.

Name	**Sequence, 5' -> 3'**
qPCR Primers	
ETS1_F	GTCAGATCTGCCTAGTCTCT
ETS1_R	AAAGAAACACACTCTGGGAA
ITS1_F	GGAGTTTTCATATCTTTGCA
ITS1_R	TGTTTGTTACCTCTGGGCCC
U3_F	GATGGGTCAAGATCATCGCG

U3_F	CTTGGTCCGCAATCCTTAGC
Actin_F	AAGGTATCATGGTCGGTATC
Actin_R	GGTGACAATACCGTCTTCAA
Northern probes	
5' ETS	GGTCTCTCTGCTGCCGG
18S	GTCTCAAAGATTAAGCCATG
ITS	TAGGACAATTAAAACCG
U3snoRNA	TAGATTCAATTTCGGTTT
SCR	TCTCAACGTATCCCATCCCA
Sanger sequencing primers	
U3_seq_primer_F	CGGGCCTCTTCGCTATTACG
U3_seq_primer_R	TCTTTTCACATACAGCGCCT
M13 primer	GTAAAACGACGGCCAGT
U3snoRNA HiSeq / NextSeq library creation and sequencing	
Custom_Read1_seq_p	TATGGTAATTGTAAACGACGGCCAGTGAATTC
R_PCR_U3bar_seq	AATGATACGGCGACCACCGAGATCTACACTATGGT AATTGTAAACGACGGCCAGTGAATTC
Index1_PCR_bar_seq	CAAGCAGAAGACGGCATACGAGATCGTGATGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCAGT TAAGAAATTGG
Index2_PCR_bar_seq	CAAGCAGAAGACGGCATACGAGATACATCGGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCAGT TAAGAAATTGG
Index3_PCR_bar_seq	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCAGT TAAGAAATTGG
Index4_PCR_bar_seq	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCAGT TAAGAAATTGG
Index5_PCR_bar_seq	CAAGCAGAAGACGGCATACGAGATCACTGTGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCAGT TAAGAAATTGG
Index6_PCR_bar_seq	CAAGCAGAAGACGGCATACGAGATATTGGCGTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCAGT TAAGAAATTGG
Adding mutations to U3snoRNA library	
EcoR1_frag_F	CATGCTACGTCCGAATTCNNNNNNNNNNNNNNNNN NNNTAATCCAATTTCTTAACTGA
EcoR1_frag_R	GTACTTCATAGGATCATTTC
Kpn1_frag_1_F	TCCTATGAAGTACGTCGADACAGATT
Kpn1_frag_2_F	TCCTATGAAGTACGTCGBCACAGATT
Kpn1_frag_3_F	TCCTATGAAGTACGTCHACACAGATT
Kpn1_frag_4_F	TCCTATGAAGTACGTDGACACAGATT
Kpn1_frag_5_F	TCCTATGAAGTACGVCGACACAGATT
Kpn1_frag_6_F	TCCTATGAAGTACHTCGACACAGATT
Kpn1_frag_R	CACATGGTACCCAGGCAGAGTTCCCCCCAG
Creating / genotyping AID strains	
Hnh_resistance_F	CGGCCGCATAGGCCACTA

Hnh_resistance_not1_R	ATAAGAATGCGGCCGCTGTTTAGCTTGCCTCGTCCC
Can1_TIR1insert_F	ATATTGGTATGATTGCCCTTGGTGGTACTATTGGTACAGGTCTTTTCATTGCCATCATCATATCGAAGTTTCA
Can1_TIR1insert_R	TTTGTGAGAAAACTGTGAAAGAGGATGTAACAGGGATGAATGTAGCCATTGTTAAAGCCTTCGAGCGTCC
Can1_check_F	CCGACGAGAGTAAATGGCGA
Can1_check_R	CCGCCAGTGGAACTTTGTAC
Imp4_degron_F	AGTGGCAGTTGAGAAGATTCATAAGGACTGCCAATAAAAAAGACTATTTGCGTACGCTGCAGGTCGAC
Imp4_degron_R	TAAGGCCTCATCGGCCTTCTATTTTAACCTTTACAAATCTATAAAAAAATATCGATGAATTCGAGCTCG
Rrp9_degron_F	CAGGTGCCAGAAATGGTATATATTCAGCTGTCATTGACCAAACAGGCTTTCGTACGCTGCAGGTCGAC
Rrp9_degron_R	CGTTCTGAGAGAGTGAAAAGGGCATGGAAAATTACGGCAATATCAGCAATATCGATGAATTCGAGCTCG
Nop56_degron_F	AAAAGAAAGATAAAAAGGAAAAGAAGGATAAAAAGAAGAAAAGTAAGGATCGTACGCTGCAGGTCGAC
Nop56_degron_R	AGAAAAGTACACAAAAAGATGGGATATACTTTATTTCGATTCATTGTTCCATCGATGAATTCGAGCTCG
Nop58_degron_F	AGAAAGAGAAGAAGGAAAAGAAGTCCAAGAAAGAGAAGAAAGAGAAGAAACGTACGCTGCAGGTCGAC
Nop58_degron_R	ATAAAAGGGAACGCGAGGGGTCACTAATTATTAAAATGTAAAATGCATCCATCGATGAATTCGAGCTCG
Imp3_degon_F	AGAAAACCTTGTTGAGATACAGAAACCAAATCGACGATTTTGATTTTTCACGTACGCTGCAGGTCGAC
Imp3_degron_R	GAAGATGATAAATCGAGTATTGATACAGAAAACGTAAGTTTGTAGTCAATATCGATGAATTCGAGCTCG
Mpp10_degon_F	GTAAAACGAAGAAGAGCAGATCAGGGCCAGATAGCACAAATATAAAACTTCGTACGCTGCAGGTCGAC
Mpp10_degon_R	TGAAGAAATAATTTTTTACCTAAATACAAACTAAGTCTTCTTCGGCGTGTATCGATGAATTCGAGCTCG
Snu13_degron_F	TCAAGACCCAAATCTACGCTGTCAAGGACAAGATTGAAACTTTATTAATTCGTACGCTGCAGGTCGAC
Snu13_degron_R	TAGTATATAGATGGTACAAAAACTATGTTGATAGATATAATCCAATCTACATCGATGAATTCGAGCTCG
Dhr1_degron_F	TCACCAGAAAGGGCTTCCAGACCATCACAGGTGAAGAGAAAGAAAAAAAACGTACGCTGCAGGTCGAC
Dhr1_degron_R	TTCCTATATGTTCTTATATAATGAAACTCTATTTCTATAATACCAGCGCAATCGATGAATTCGAGCTCG
Emw1_degron_F	GGGATGACTCTCCTGGTTGGGATAGGCTAGTGGAAGCAAGAAGCCAAATACGTACGCTGCAGGTCGAC
Emw1_degron_R	TTTATACAAACATTACGTAATATACACAGATTATACATGAAAGGTGCTCTATCGATGAATTCGAGCTCG
Nop1_degron_F	GAGACCATTGTATCGTCGTTGGTAGATACATGAGAAGCGGTTTGAAGAAACGTACGCTGCAGGTCGAC
Nop1_degron_R	AATAATTGAGCAAGCTACCAGAAACTTTTGACTAATTTCCTTTATTCAACATCGATGAATTCGAGCTCG

Imp4_check_F	GAGATTGCCGAAGTTGGTCC
Imp4_check_R	ACCTCTTCCCTTCGTTCGTT
Rrp9_check_F	TGCCAAAGAACACAGACTCG
Rrp9_check_R	GCTTTGCGTTCTGAGAGAGT
Nop56_check_F	GGAAGAAACCAAGGAAAAGGAAT
Nop56_check_R	GGCGAGTGAAAGAGGCATAA
Nop58_check_F	GACGATGATGAAGATTCCAAGGA
Nop58_check_R	AACGCGAGGGGTCACTAATT
Imp3_check_F	TCAACGATCCAGCCTACCTC
Imp3_check_R	TGTATTTGCTGCCTCGTTGC
Mpp10_check_F	AAGAAGAGCAGATCAGGGCC
Mpp10_check_R	AGGGGTTAATGACATTACAAGGT
Snu13_check_F	TACCACCAACGATGCTTCTG
Snu13_check_R	CGTCTGCACTAGTGGAAAGC
Dhr1_check_F	GATTTGGATCCACAGTTGACAAC
Dhr1_check_R	GCTCCATTATACAAGCCTCCTG
Emw1_check_F	TCTGGTTTGCAAAGACTGGA
Emw1_check_R	CACCCAGCCAAGGAGATAAAAT
Nop1_check_F	CTACTGTAGACGCGGAAACC
Nop1_check_R	AGGAAGTGAAACTGAATGGGG
Genotype_outsideMAT	AGTCACATCAAGATCGTTTATGG
Genotype_alpha	GCACGGAATATGGGACTACTTCG
Genotype_a	ACTCCACTTCAAGTAAGAGTTTG

2.5 Plasmids

Name	Description	Source
U3snoRNA library		
pU3-wt	Contains wild type U3snoRNA expressed by native promoter, ADE2 marker	(Puchta, et al., 2016)
pU3-small_library (pool)	pU3-wt plasmid but with mutant library replacing WT sequence	(Puchta, et al., 2016)
pU3-new_mutants (pool)	pU3-wt plasmid but with new mutants replacing WT sequence	(Puchta, et al., 2016)
PU3-small_library+new_mutants (pool)	pU3-wt plasmid but with mutant library and new mutants replacing WT sequence	This thesis

Auxin inducible degron system		
pMK68 (BYP7425)	Contains 1x mini auxin inducible degron sequence, kanamycin cassette	National BioResource Project, Japan (Masato Kanemaki)
pMK68 (BYP7425) +3xFLAG	Contains 1x mini auxin inducible degron sequence + 3FLAG tag, kanamycin cassette	This thesis
pMK151 (BYP7430)	Contains 3x mini auxin inducible degron sequence, kanamycin cassette	National BioResource Project, Japan (Masato Kanemaki)
pMK200 (BYP7569)	Contains ADH1-689-OsTIR1, kanamycin cassette	National BioResource Project, Japan (Masato Kanemaki)
pSM409 (BYP6851)	Contains auxin inducible degron, hygromycin B cassette	National BioResource Project, Japan (Masato Kanemaki)
pMK200 (BYP7569 TIR1) + hph	Contains ADH1-689-OsTIR1, hygromycin B cassette	This thesis

2.6 Strains and cell lines

Strains were created by using primers with overhangs containing homology for the different sites of integration, using this primers to create a PCR product from the plasmid containing the auxin inducible degron sequence / TIR1 sequence.

Name	Genotype	Source
Yeast strain		
D343	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2	(Puchta, et al., 2016)
D343-TIR1	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized)	This thesis

D343-Nop1-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), NOP1::NOP1-miniAID-KanMX	This thesis
D343-Imp4-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), IMP4::IMP4-miniAID-KanMX	This thesis
D343-Rrp9-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), RRP9::RRP9-miniAID-KanMX	This thesis
D343-Nop56-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), NOP56::NOP56-miniAID-KanMX	This thesis
D343-Nop58-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), NOP58::NOP58-miniAID-KanMX	This thesis
D343-Mpp10-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), MPP10::MPP10-miniAID-KanMX	This thesis
D343-Imp3-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), IMP3::IMP3-miniAID-KanMX	This thesis
D343-Dhr1-AID-3FLAG	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), DHR1::DHR1-miniAID-3FLAG-KanMX	This thesis

D343-Emw1-AID	leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), EMW1::EMW1-miniAID-KanMX	This thesis
D343-Nop1-AID_haploid	MAT α, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), NOP1::NOP1-miniAID-KanMX	This thesis
D343-Imp4-AID_haploid	MAT a, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), IMP4::IMP4-miniAID-KanMX	This thesis
D343-Rrp9-AID_haploid	MAT a, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), RRP9::RRP9-miniAID-KanMX	This thesis
D343-Nop56-AID_haploid	MAT α, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), NOP56::NOP56-miniAID-KanMX	This thesis
D343-Nop58-AID_haploid	MAT a, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), NOP58::NOP58-miniAID-KanMX	This thesis
D343-Mpp10-AID_haploid	MAT α, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), MPP10::MPP10-miniAID-KanMX	This thesis
D343-Imp3-AID_haploid	MAT a, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), IMP3::IMP3-miniAID-KanMX	This thesis

D343-Dhr1-AID-3FLAG_haploid	MAT a, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), DHR1::DHR1-miniAID-3FLAG-KanMX	This thesis
D343-Emw1-AID_haploid	MAT α, leu2, ura3, ade2, can1, his1, his3, trp1, U3aΔ, UASGAL:U3A::URA3, U3B::LEU2, can1::P_{AHD1}-OsTIR1 (codon optimized), EMW1::EMW1-miniAID-KanMX	This thesis

2.7 Experimental Protocols

2.7.1 General culturing and preservation of yeast cultures

Yeast liquid cultures were grown at 30°C in flasks and shaken at 250rpm with a rotary shaker. The media was kept under 20% of the flask volume. When grown on plates yeast were grown at 30°C, and kept at 4°C for storage for up to one month. Plates had a volume of 25 ml, and were poured a period of time after autoclaving when still hot enough to pour, and left to cool and set. For long-term storage a saturated culture was diluted 1:1 with 40% glycerol and stored at -80°C. Yeast were grown in YPGalactose for most purposes unless specified.

To obtain an overnight culture media was inoculated from a yeast plate colony, and left to grow overnight. To obtain log-phase cultures an overnight culture was diluted to 0.2 OD_{600} and allowed to grow for around ~6 hours until at least 2 divisions had occurred (>0.8 OD_{600}).

When pelleting cells, yeast were centrifuged at 4°C at 3,000 x g for 3 minutes. For harvesting pellets the yeast were then resuspended in 10% culture volume of H_20, centrifuged at 3,000 x g for 3 minutes, the supernatant removed, and the pellets stored at -20°C.

2.7.2 Yeast transformation

For the transformation of yeast I utilised the high-efficiency version of the lithium acetate/single-stranded carrier DNA/PEG method of transformation of *Saccharomyces cerevisiae* described in (Gietz & Schiestl, 2007). I used this

protocol for both the transformation of oligos intended to be integrated into the genome, and for plasmids transformations.

A 5 ml overnight culture was setup the day before the transformation, with a 250 ml flask and 150 ml of YPGalactose x2 also placed at 30°C in preparation. The next day the overnight culture of yeast were diluted to 0.5 OD_{600} in 150 ml of YPGalactose x2 media. This culture was then incubated at 30°C at 250rpm until the cell titer is at least 2.5 OD_{600}, ensuring at least two divisions have occurred, taking around 6 hours. Immediately prior to this 1 ml of salmon sperm carrier DNA (2mg/ml) is boiled for 10 mins, and then kept on ice. The cells are harvested by centrifugation at 4°C at 3000 x g for 3 min. The supernatant is removed, the pellet is then resuspended in 20 ml of water, centrifuged again, supernatant removed, resupsended in 1ml of water and transferred to a 1.5 ml tube. From this 1 ml of suspension, ten 100 µl aliquots can be transferred into ten1.5 ml tube, centrifuged, and the supernatant removed, and kept on ice for ten transformations.

The following transformation mix is then made up:

	X1
PEG 50%	240 µl
LiAc 1.0M	36 µl
SS Carrier DNA (2mg/ml)	50 µl
DNA (oligo or plasmid)	34 µl
total	360 µl

The transformation mix is then added to the yeast pellet, making sure it is fully resuspended. This is then left on ice for 30 mins. After this the tube(s) are placed in a floating rack and incubated at 42°C for 35 minutes (this was found to be the best time for the D343 strain). The tubes are then span at 10,000 x g for 30 seconds, and the transformation mix is removed leaving the yeast pellet.

If performing a low-throughput transformation of plasmids or oligos, the pellet is then resuspended in 0.5 ml of sterile water. This can then be plated out at appropriate dilutions onto the YPGalactose selection plate, and left for 3-4 days for colonies to grow. Some selection markers such as Geneticin (G418) or Hygromycin have phenotypic lag and require time for the yeast to express the resistance marker. This is achieved by resuspending in 0.5 ml of YPGalactose instead of water and left at 30C for >3 hours to allow for the yeast to recover before plating on selection plates.

2.7.3 AID strain creation transformations

For integration of TIR1 to the Can1 locus I PCR amplified the TIR1 sequence along with hph resistance cassette from the pMK200_hph plasmid (**Chapter 3: Augmentations**), using primers with 50 bp overhangs with homology for the Can1 site (**2.4 Oligonucleotides**). The homology arms meant that after transformation homologous recombination would integrate TIR1 and the hnh resistance cassette at the Can1 site.

For the creation of the AID strains I followed the protocol from (Nishimura & Kanemaki, 2014) for C-terminal tagging. The plasmids used in these protocols are available from the National BioResource Project, Japan. I used

the pMK68 (1x mini AID, kanamycin cassette) to provide a template for the mini AID tags. I PCR amplified a sequence from the pMK68 plasmid, containing the AID sequence and a stop codon. The primers had overhangs with 50 bp homology arms for the genomic sites of the particular protein of interest. The homology sites being either side of the genomic stop codon ensured the genomic stop codon was replaced with the AID sequence and new stop codon, making the degron part of the protein of interest's open reading frame.

To create the oligos with homology arms I purified the product of four 50µl Accuprime PCR reactions using the QIAquick PCR Purification Kit and then used 9 µg of this product in a x1 transformation. The strains were allowed to recover >3 hours in YPGalactose at 30°C to account for phenotypic lag of the antibiotic resistance, and then plated on the appropriate antibiotic plates (Hygromycin B for Tir1, and Geneticin (G418) for the AID strains).

2.7.4 Yeast competition experiment

The competition experiment is the wet lab portion of the high-throughput fitness assay. For transforming the U3 mutant plasmid library the same high-efficiency version of the lithium acetate/single-stranded carrier DNA/PEG method of transformation was used. Everything however was scaled 30 fold, including number of yeast and transformation mix: This was to ensure the entire library was transformed and thus represented in the assay, aiming for at least 400,000 transformants.

	X30 (fitness assay)
PEG 50%	7.2 ml
LiAc 1.0M	1.08 ml
SS Carrier DNA (2mg/ml)	1.5 ml
DNA (oligo or plasmid)	1.02 ml
total	10.8 ml

After heatshock the transformation mix was transferred first to 50ml of CSM -ADE Galactose media, with 100µl plated out on a CSM -ADE Galactose plate, then to 1L of CSM -ADE Galactose media, with 100µl plated out on a CSM -ADE Galactose plate. The two plates are used to estimate the efficiency of transformation by counting colonies after incubating at 30°C for two days. The 1L liquid culture was grown at 30°C at 230 RPM until it reached OD_{600} >2. This was split to OD_{600} 0.2 (~2 x 10^9 cells), and grown to OD_{600} 1 to allow a culture of majority transformed yeast. This population was termed 'G0', and represented the first time point that would be sequenced. Time points were collected by centrifuging 50ml of culture at 3000 x g for 3 minutes, removing the supernatant, resuspending in 1ml of sterile water, centrifuging at 3000 x g for 3 minutes, removing the supernatant, and storing the pellet at -20°C.

After G0 the culture was split to OD_{600} 0.1 in 500ml of CSM -ADE Glucose, to downregulate the genomic U3 snoRNA expression from a galactose inducible promoter. The culture was split to OD_{600} 0.1 every 12 hours / 24 hours (depending on experiment) to maintain log growth (<1 OD_{600}). Time points were collected every time the culture was split, with the time points chosen for sequencing being determined after the experiment was complete.

2.7.5 Auxin

The auxin 1-Naphthaleneacetic acid (NAA) was added to the media in different concentrations (100 µM -1000 µM) for use in the auxin inducible degron system. Before adding to media, the NAA was first dissolved in ethanol to create a 1M stock, with the required concentration being diluted 1 in 25 in 40% ethanol before adding to media to facilitate it being dissolved more efficiently.

2.7.6 Sporylation and isolation of haploids

In the presence of a non-fermentable carbon source and nitrogen starvation diploid yeast will package into haploid spores (Freese EB, 1982). A 5ml overnight culture of yeast were grown, and diluted to 0.2 OD_{600} in the morning. This was allowed to grow for six hours at 30°C until log phase was achieved (OD_{600} >0.8). Cells were centrifuged for 3 minutes at 3000 x g, resuspended in 1 ml of 1% potassium acetate, centrifuged for 3 minutes at 3000 x g and resuspended in 2 ml of 1% potassium acetate supplemented with 5 mg/L of adenine. This supplementation of adenine at a quarter of the amount found in SD media is due to the D343 strain's adenine auxotrophy. The cultures were then incubated for >4 days in a culture tube at room temperature on a laboratory roller.

The level of sporulation can then be observed by microscopy by removing 10µl and putting it on a glass slide with cover slip atop. Usually find >25% tetrads after 5 days with the D343 strain. The yeast are then centrifuged at 3000 x g for 3 minutes, resuspended in 1ml of KPi buffer A, centrifuged at 3000 x g for 3 minutes and resuspended in 250 µl of KPi buffer B (Buffer A + zymolyase).

This is then kept at 37°C for 2 hours to allow the zymolyase to digest the membrane sac of the asci tetrads. The yeast are then centrifuged at 3000 x g for 3 minutes, and resuspended in 400 µl of sterile water. This is then plated on two selection plates, 300µl on one, 100µl on the other.

When selecting for haploid AID strains, they were plated on both hygromycin (Tir1 selective marker) and G418 (degron selective marker). This meant any resulting colonies were either haploid containing both TIR1 and the degron, or diploid. These could be distinguished by choosing smaller colonies that were presumed haploid, and doing colony PCR for the MAT locus to confirm ploidy, and then for TIR1 and degron to confirm genotype.

2.7.7 Yeast DNA/RNA phenol:chloroform extraction and ethanol precipitation

To isolate high quality DNA or RNA from yeast cultures I used a phenol:chloroform method. This was carried out underneath a hood. For a 5 ml log culture of yeast, the yeast are centrifuged for 3000 x g for 3 minutes. The supernatant is removed, and resuspended in 400µl of AE buffer and 40µl of 10% SDS. 400µl of phenol is then added (pH 4.2 for RNA, pH 6.6 for DNA), with the lids securely sealed with parafilm. This is then placed in a thermomixer at 65°C, at 6000 rpm for 5 minutes. The sample is placed at -80°C for 10 minutes. It is then centrifuged at 20,000 x g for 10 minutes. 300µl of the upper water phase is transferred to a new tube, with 300µl of chloroform being added to it and then vortexed. This is centrifuged for 20,000 x g for 10 minutes. 250µl of the upper phase is then transferred to a new tube.

The DNA/RNA is then precipitated using ethanol. 0.1 volume of 3M pH 5.3 NaAc is added, with 3 volumes of ice cold ethanol. This is mixed by pipetting. This is then stored at -80°C for >2 hours or overnight at -20°C. The sample is then centrifuged at 20,000 x g for 30 minutes at 4°C, with the supernatant subsequently being removed. The pellet is then washed 0.7 volume of 70% ethanol, and centrifuged at 20,000 x g for 30 minutes at 4°C. The supernatant is removed, and the pellet is left to air dry. The pellet is then resuspended in the desired amount of sterile water.

2.7.8 Yeast colony PCR

When high quality DNA was not necessary, but instead needed to genotype many clones, I used a faster colony PCR method of DNA extraction. Using a sterile tip, a just visible amount of yeast were taken from a colony and resuspended in 20µl of 20 mM NaOH. This was then incubated at 95°C for 45 minutes. This is then centrifuged at 3000 x g for 3 minutes to pellet the cells, and 2µl of the supernatant is used in 25µl PCR reaction. This PCR reaction uses a 100µM concentration of primers (10 fold more than usual).

2.7.9 PCR / bacterial culture DNA purification

When purifying DNA from PCR reactions I used the QIAquick PCR Purification Kit or MinElute PCR Purification Kit, and followed the supplied protocol, eluting in EB buffer.

When purifying plasmids from bacterial cultures I used the QIAprep Spin Miniprep Kit or QIAGEN Plasmid Plus Maxi Kit, and followed the supplied protocol, eluting in EB buffer.

2.7.10 Protein extraction and western blotting

Yeast were centrifuged for 3000 x g for 3 minutes. Resuspended in 100µl of water, and then added 100µl of 0.2M NaOH. Incubated at room temperature for 5 minutes. Centrifuged for 3000 x g for 3 minutes. Removed supernatant and resuspended cells in 50µl of NuPAGE LDS Sample Buffer (4X). Heated tube to 70°C for 3 minutes, and then centrifuged for 3000 x g for 3 minutes. Aliquoted supernatant into new tube, and discarded pellet. Loaded 15-20µl of supernatant into each lane of a NuPAGE 10% Bis-Tris Protein Gel. Ran protein gel at 200V with NuPAGE MOPS SDS Running buffer until dye front reached bottom of the gel.

Followed the manufactures protocol for using the iBlot 2 Transfer Stacks, to transfer protein to nitrocellulose membrane.

Made 5% milk with dry milk powder in TBS-T. Rinsed membrane with TBS-T and then blocked membrane in 5% milk TBS-T overnight on a rocker at 4°C. In the morning washed the membrane for 5 minutes on a rocker in TBS-T four times. Incubated in 5% milk TBS-T with primary antibody at correct dilution overnight on a rocker at 4°C. In the morning washed the membrane for 5 minutes on a rocker in TBS-T four times. Incubated in 5% milk TBS-T with secondary antibody at correct dilution for 1 hour on a rocker at room

temperature. Washed the membrane for 5 minutes on a rocker in TBS-T four times.

Followed manufactures protocol to use Pierce Supersignal Femto kit to develop membrane. Then used ImageQuant™ LAS 4000 to visualise blot.

2.7.11 Determining nucleic acid concentrations and quality

The concentration and quality of nucleic acids within a sample were spectrophotometrically assessed by optical density at 260nm using a NanoDrop (Thermo Scientific). When submitting libraries for high-throughput sequencing and thus a higher sensitivity was required, I used an Agilent Bioanalyser 2100.

2.7.12 Polymerase chain reaction parameters

PCR amplification was performed using either Phusion High-Fidelity PCR Master Mix with HF Buffer or AccuPrime Pfx SuperMix, with it being specified for each experiment. I followed the manufacturer's protocols, with the cycling conditions as follows:

Phusion High-Fidelity PCR Master Mix with HF Buffer, standard PCR amplifications, conditions dependent on particular PCR:

Stage	Temperature	Time
Initial denaturation	98°C	30 seconds
Amplification (25-35 cycles)	98°C	5-10 seconds
	45-72°C	10-30 seconds
	72°C	15 seconds per kb of plasmid DNA / 30 seconds per kb of genomic DNA
Final Extension	72°C	5-10 minutes
Hold	4°C	-

Phusion High-Fidelity PCR Master Mix with HF Buffer, library preparation PCR amplifications:

Stage	Temperature	Time
Initial denaturation	98°C	30 seconds
Amplification (25 cycles)	98°C	10 seconds
	45-72°C	10 seconds
	72°C	10 seconds
Final Extension	72°C	10 seconds
Hold	4°C	-

AccuPrime Pfx SuperMix, standard PCR amplifications:

Stage	Temperature	Time
Initial denaturation	95°C	5 minutes
Amplification (25-35 cycles)	95°C	15 seconds
	55-65°C	30 seconds
	68°C	1 minute per kb
Hold	4°C	-

2.7.13 Agarose gel electrophoresis

To visualise and diagnose DNA fragments, agarose gel electrophoresis was used. A 100 ml 1% gel was prepared using 1g/100ml of Agarose in 1xTBE, with 2ul of Ethidiumbromide.

2.7.14 Sanger sequencing

Sanger sequencing was performed by the IGMM HGU technical services. I then aligned and analysed the sequences via the SnapGene program by GSL Biotech.

2.7.15 Library Creation

The construction involved two PCR reactions, with six overlapping oligos (~80 nucleotides in length) that together constituted the 333 nucleotide U3 snoRNA sequence plus flanking regions (**Figure 7A**). The oligonucleotides could be

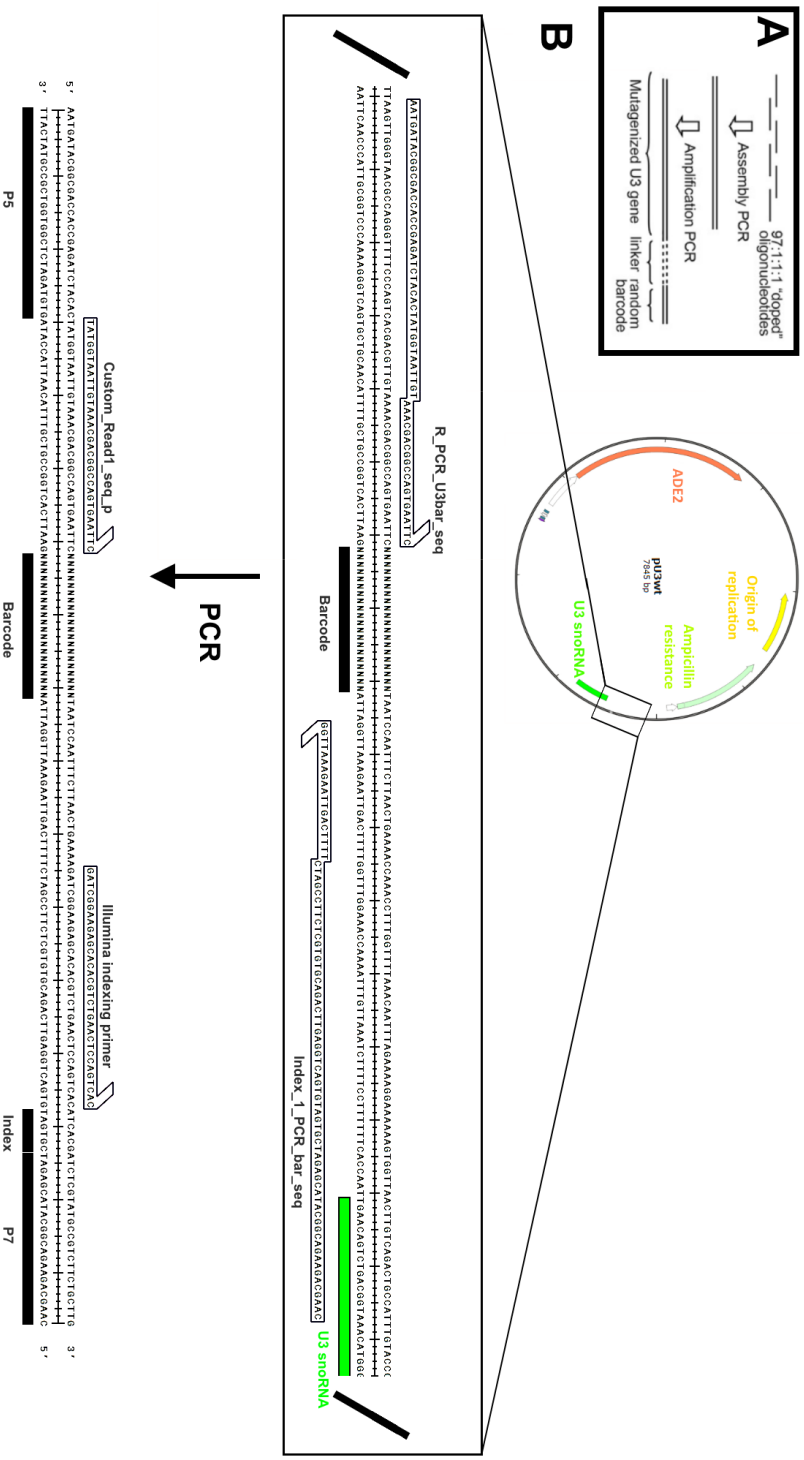


Figure 7, Library creation and barcode sequencing strategy

A. Schematic of the PCRs used for mutant library creation. (From (Puchta, et al., 2016)).
B. Schematic of the pU3wt plasmid, and where the R_PCR_U3bar_seq and Index*_PCR_bar_seq primers anneal to produce an oligo for Illumina sequencing. On this oligo the location of the custom sequencing primer (Custom_Read1_seq_p) and Illumina indexing primer used in Illumina sequencing are shown.

“doped” when ordered, which meant at each position there was a 97% chance of a WT nucleotide, and a 1%:1%:1% mix of the alternative bases. This is what allowed the mutagenesis approach to be unbiased and comprehensive.

When creating the “big” library six doped oligos were used, and for the “small” library six separate PCR reactions were performed, with only one of the oligos being doped each time. In this thesis I used the small library, as this has ~1% mutation rate rather than the ~3% of the big library. As only one of the six oligos was mutated per variant, certain mutational combinations should be impossible in the “small” library. Mutations should only be found in widows of nucleotide positions 6-88, 69-152, 132-215, 195-278, 259-333 and 321-333. The “big” library was utilised to study intramolecular epistatic interactions, as mutation combinations could be distributed anywhere along the breadth of U3 snoRNA.

After the six oligos were assembled a second PCR was used to amplify the sequence, with primers overhangs containing a 20 nucleotide random sequence used to assign a random barcode to each variant. Conveniently, the first six nucleotides of the U3 snoRNA sequence constitute a SalI restriction site. This region was consequently not “doped” as the SalI site along with an EcoRI site on the other side of the sequence was used to clone into a vector (pU3wt) which contained a native U3 promoter. The library was then transformed into DH5α *E. coli*, with subsequent colonies being pooled and plasmids extracted to obtain the U3 snoRNA plasmid mutant library.

A 400 nucleotide fragment of the plasmid library including the barcode, linker sequence and U3 snoRNA sequence was sequenced using 300-nt paired-end

MiSeq Illumina sequencing to associate molecular barcodes with their respective U3 snoRNA sequence, detailed in (Puchta, et al., 2016). This facilitated sequencing only the barcode sequence in future experiments with a custom primer, as the barcodes and associated U3 snoRNA sequence were known. The MiSeq data was also used to check the complexity of the library and identify reliable variants with the following filtering parameters. First barcodes that were too similar were rejected to avoid miss assignment. A mutation in the U3 snoRNA sequence had to appear in at least 80% of the assigned reads and have a minimal local coverage of 80 reads, with anything less being rejected. This threshold was chosen due to the large reduction in noise when comparing the fitness values of wild type variants. Additionally any variants with mutations in the linker sequence between the U3 snoRNA sequence and barcode were rejected. This resulted in ~41,000 total identified barcodes from the “small” library being reduced to ~23,000 high confidence barcodes, including ~3000 wild type variants. In the “small” library mutations should only be found in widows of nucleotide positions 6-88, 69-152, 132-215, 195-278, 259-333 and 321-333. This however was found to not be the case all of the time, with mutations being found across windows, with these mutations perhaps occurring during oligo synthesis or library creation.

2.7.16 Preparing samples for MiSeq, HiSeq, NextSeq

Plasmid DNA was extracted from yeast collected at timepoints during the competition experiment. This was achieved using a modified version of the QIAprep Spin Miniprep Kit protocol. Using OD_{600} measurements (a 1 ml culture at 0.1 OD_{600} is ~1x10^6 yeast), I aimed to use ~10 x10^7 yeast in each extraction, in line with the amount of bacteria recommended for the protocol. The yeast were centrifuged at 3000 x g for 3 minutes, resuspended in 1ml of KPi buffer A, centrifuged at 3000 x g for 3 minutes and resuspended in 250 µl of KPi

buffer B (Buffer A + zymolyase). This was then kept at 37°C for 2 hours to allow the zymolyase to digest the cell wall. The yeast are then centrifuged at 3000 x g for 3 minutes, and resuspended in 250µl of buffer P1 from the QIAprep Spin Miniprep Kit. The kit protocol is then followed the same proceeding the addition of buffer P1, with the zymolyase digestion of the cell wall making the yeast now suitable to use in the kit.

16ng of template (~1.5 x 10^9 molecules) was amplified with the R_PCR_U3bar_seq and Index*_PCR_bar_seq primers in eight 50µl Phusion High-Fidelity PCR reactions per timepoint. This amplified the barcode sequence while attaching sequencing adapters, with each time point having a different indexing primer (**Figure 7B**). PCR bias should be limited as many genotypes are represented by multiple barcode sequences. The eight PCR reactions of each time point were pooled to 25µl using the QIAquick PCR Purification Kit. The appropriate band (168bp) was then isolated by 2% E-Gel SizeSelect agarose gel electrophoresis, and quantified by Bioanalyser. An equimolar pool of each time point ready for sequencing was then created using the Bioanalyser readings.

2.7.17 Calculating log fitness values

The population size of an experiment involving continuous growth and overlapping generations can be represented by '$N_t = N_0 \exp(m\ t)$'. Where N_t is the number of individuals at time t, the initial number of individuals is N_0, and the exponential growth rate is m ("Malthusian parameter"). The Malthusian parameter of a population can be thought to be the logarithm of fitness. This is the basis behind the calculation of log fitness in the high throughput fitness assay, with each barcode (genotype/population) being assigned a log fitness

value. These values are generated using a Poisson regression approach with exponentially decaying mean. In the following formula, barcode counts are l with time being t, modelled as Poisson random variables. The unknown log fitness of barcode l, b_t and b_{l0} are represented by λ_l, used to as normalisation factors to account for different initial counts and library sizes. Gaussian priors were placed over the λ_l, b_t and b_{l0} variables, and obtained Bayesian posterior estimates using the Expectation-Propagation approximation. These calculations were done in collaboration with Guido Sanguinetti from The School of Informatics at The University of Edinburgh.

$$n_l(t) \sim Po(m_l^t)$$

$$m_l^t = \frac{\exp(\lambda_l t + b_{l0})}{b_t}$$

These calculations provide a log fitness value for each barcode, generated from the read counts for each barcode. This is then normalised by subtracting the median log fitness of the ~3000 wild type genotype barcodes.

As well as obtaining a log fitness value for each of the genotypes, the results can be aggregated to look at positional effects. The fitness effect for each U3 snoRNA position are can be represented as:

f_i

Average log fitness effect of mutations at position *i* in a wild-type genetic background (f_i), calculated as the mean log fitness of variants that had a single substitution or deletion at position *i*.

p_i

Aggregate log fitness effect of mutations at position i in all genetic backgrounds (p_i), calculated as the mean log fitness of variants that had a substitution or deletion at position *i*, minus the mean log fitness of variants that had no substitution or deletion at position *i*.

2.7.18 High-throughput sequencing

High-throughput sequencing was carried out by Edinburgh Genomics (HiSeq) and The Edinburgh Clinical Research Facility (NextSeq). I then processed and analysed the sequencing data using the University of Edinburgh's Eddie3 compute cluster.

2.7.19 DNAse treatment

DNAse treatment of RNA was carried out using the Turbo DNAse kit, following the supplied protocol. 9µg of RNA, quantified by Nanodrop, was used in the 'Routine DNase treatment' procedure.

The DNA/RNA is then precipitated using ethanol. 0.1 volume of 3M pH 5.3 NaAc is added, with 3 volumes of ice cold ethanol. This is mixed by pipetting. This is then stored at -80°C for >2 hours or overnight at -20°C. The sample is then centrifuged at 20,000 x g for 30 minutes at 4°C, with the supernatant subsequently being removed. The pellet is then washed 0.7 volume of 70% ethanol, and centrifuged at 20,000 x g for 30 minutes at 4°C. The supernatant is removed, and the pellet is left to air dry. The pellet is then resuspended in the desired amount of sterile water.

2.7.20 cDNA creation

The generation of cDNA from RNA was achieved via the Superscript III reverse transcriptase. 1µg of DNAse treated RNA, quantified by nanodrop, was used in each reaction with random primers, following the supplied protocol.

2.7.21 Quantitative polymerase chain reaction

RT-qPCR reactions were carried out using LightCycler480 SYBR Green I Master Mix, on the Roche lightcycler 480 machine. All RT-qPCR reactions

used 10µl of LightCycler 480 SYBR Green master mix x2, 2µl of cDNA (diluted according to primer standard curve results), 0.6µl of 10mM gene specific forward primer, 0.6µl of 10mM gene specific reverse primer and 6.8µl of sterile water. The reactions took place in a 96 well plate. The ct values generated were then interpreted using the Pfaffl method (Pfaffl, 2001).

Stage	Temperature	Time	Ramp Rate
Pre-incubation	95°C	5 minutes	4.4
Amplification (x45 cycles)	95°C	10 seconds	4.4
	55°C	7 seconds	2.2
	70°C	10 seconds	4.4
Melting curve	95°C	5 seconds	4.4
	65°C	1 minute	2.2

2.7.22 Northern blot

A 1% agarose NBC denature gel is ran out (Agarose 1g, 10x NBC 10ml, Formaldehyde (37%) 2.5ml, H_20 90ml).

2µg of RNA in 5µl is added to 15µl of sample preparation buffer (10x NBC 2µl, formaldehyde (37%) 3µl, formamide 10µl). This is incubated for 5 minutes at 65°C. 2µl of RNA loading dye is added, and then ran in the 1% agarose NBC denature gel for 4.5 hours. The RNA is then transferred to a membrane by a passive, slightly alkaline, downward elution with 10x SSC.

The 5' of the oligo probes are then labelled. (2µl of probe DNA (10µM), 2µl of PNK buffer 10x, 1µl of PNK, 2µl of gammaP^{32}-ATP (10µci/µl), 13µl of H_20).

This is incubated for 30 minutes at 37°C, and the placed in an ice bath. 80µl of H_20 is then added to 20µl of labelled probe, and added to a quckspin column, and centrifuged for 1000g for 4 minutes.

The membrane is pre-hybridised by placing it inside a tube with 20ml of PerfectHyb Plus Hybridization buffer which covers the membrane, and then incubated at 37°C for 30 minutes rotating. 40µl of the processed probe is then added to the Perfecthybe buffer, and left overnight at 37°C rotating. The next morning store Perfecthybe buffer with probe at -20°C, and wash membrane with 6xSSC for 10 minutes at 37°C, then wash with low stringency buffer (2xSSC, 0.1% SDS) for 10 minutes at 37°C, then wash with 6xSSC for 10 minutes at 37°C. Then develop the blot using X-ray film.

2.7.23 Bacterial transformation

The DH5α strain of bacteria was used when transforming. Competent cells are taken out of -80°C storage and thawed on ice for 30 minutes. 1-5µl of DNA (10pg to 100ng) is added to 20-50µl of competent cells in a 1.5ml tube, and left on ice for 30 minutes. To heat shock the bacteria the tube is placed into a 42°C water bath for 45 seconds. The tube is placed back on ice for 2 minutes. The contents of the tube are added to 300µl of SOC media to recover at 37°C for 45 minutes. 100µl aliquots of the SOC media are plated on the appropriate antibiotic plate. The plate is incubated at 37°C overnight, with colonies appearing the following morning.

3 HIGH-THROUGHPUT ASSAY AUGMENTATIONS REQUIRED TO STUDY GENETIC INTERACTIONS

3.1 Experimental system: High-throughput fitness assay

Using a high-throughput fitness assay, the individual impact of mutations as well as the epistatic interactions between them were elucidated in the fitness landscape of U3 snoRNA (Puchta, et al., 2016). This revealed the landscape at single nucleotide resolution, observing fitness heterogeneity depending on the substitution. This was achieved using a plasmid library of ~60,000 U3 snoRNA mutant variants created by saturation mutagenesis, with mutations distributed uniformly across the entire RNA (**Figure 8C**). All 981 possible single mutations and 99.4% of double mutations (52,981) were represented in the library, along with many multiple mutations. When creating the libraries, which is detailed in the Materials and Methods section of this thesis, a “small” and “large” library was created. In this thesis I used the “small” library in which mutations occur in ~80bp windows, resulting in a ~1% mutation rate rather than the ~3% of the big library. The “big” library was utilised to study intramolecular epistatic interactions, as mutation combinations could be distributed anywhere along the breadth of U3 snoRNA.

The principle of the assay relies on a yeast strain (D343), which contains a single genomic copy of U3 snoRNA under a galactose-inducible GAL promoter. When grown on glucose as the carbon source instead of galactose, genomic U3 snoRNA expression is strongly repressed. This can be rescued by transforming a U3 snoRNA mutant plasmid library, with each yeast in the population expressing a different U3 snoRNA genotype. Sequencing at successive time points allows the reproductive success relative to wild-type genotypes to be quantified by copy number. Using fitness (growth relative to a reference) as the phenotypic readout allows a quantitative measurement. The experiment is maintained to allow the yeast to achieve at least 40 generations,

allowing fitness effects to manifest. Genotypes that have a large impact upon U3's function subsequently have a large growth defect, while neutral mutations increase in number as they outcompete the population. The parallel nature with a large pool of variants all measured in identical conditions greatly reduces noise. The competition experiment methodology and the distribution of the U3 snoRNA mutant plasmid library are illustrated in **Figure 8**, and described in further detail in the Materials and Methods section of this thesis.

The high-throughput nature allows the experiment to look at a large breadth of mutations with high precision, quantifying small effects at single nucleotide resolution. A unique 20 nucleotide molecular barcode on each plasmid is used as the sequencing template, as a single sequencing read would not encompass the entire U3 snoRNA sequence. The mutant variants are quantified with the logarithm of relative fitness, by fitting exponential decay curves to the barcode read counts at each time point. These log fitness values are ultimately used to represent the impact of mutations. The results were reproducible, with an R^2 value of 0.98 between replicates.

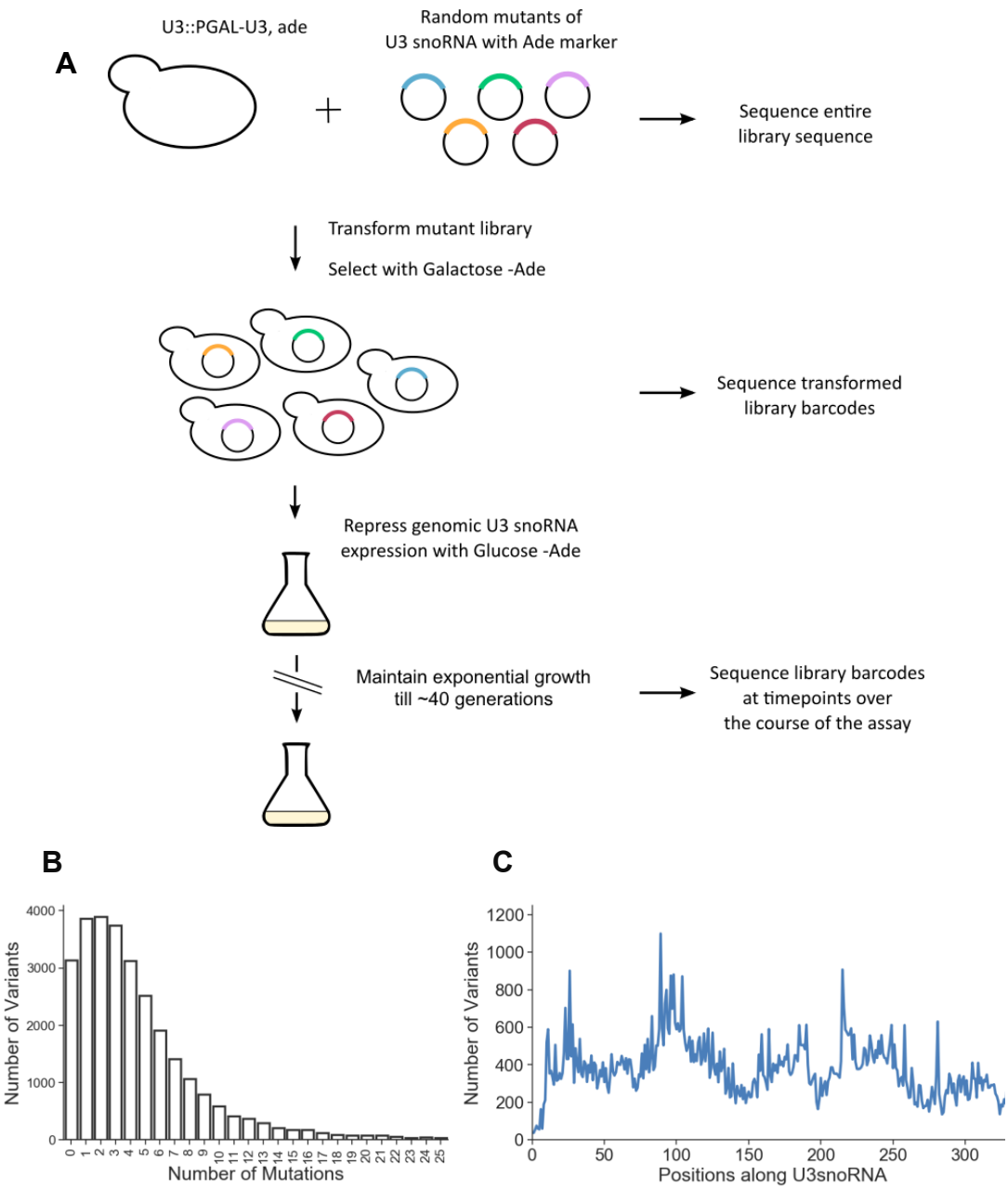


Figure 8, The competition experiment

A. The U3 snoRNA mutant plasmid library is transformed into the D343 strain, selected for with adenine auxotrophy. The genomic copy of U3 snoRNA is then downregulated by culturing in glucose, allowing the yeast to compete, forced to rely on the expressed mutant U3 snoRNA variant. Sequencing at time points subsequent to culturing in glucose allows the competing population to be quantified.
B. A histogram showing the distribution of mutation number per variant in the “small” library. Variants with 0 mutations represent the wild type variants in the library.
C. Line chart illustrating the number of variants in the “small” library that have a mutation at a position, showing the large breadth of coverage.

As well as obtaining a log fitness value for each of the genotypes, the results can be aggregated to look at positional effects. The fitness effect for each U3 snoRNA position are can be represented as:

f_i

Average log fitness effect of mutations at position *i* in a wild-type genetic background (f_i), calculated as the mean log fitness of variants that had a single substitution or deletion at position *i*.

p_i

Aggregate log fitness effect of mutations at position i in all genetic backgrounds (p_i), calculated as the mean log fitness of variants that had a substitution or deletion at position *i*, minus the mean log fitness of variants that had no substitution or deletion at position *i*.

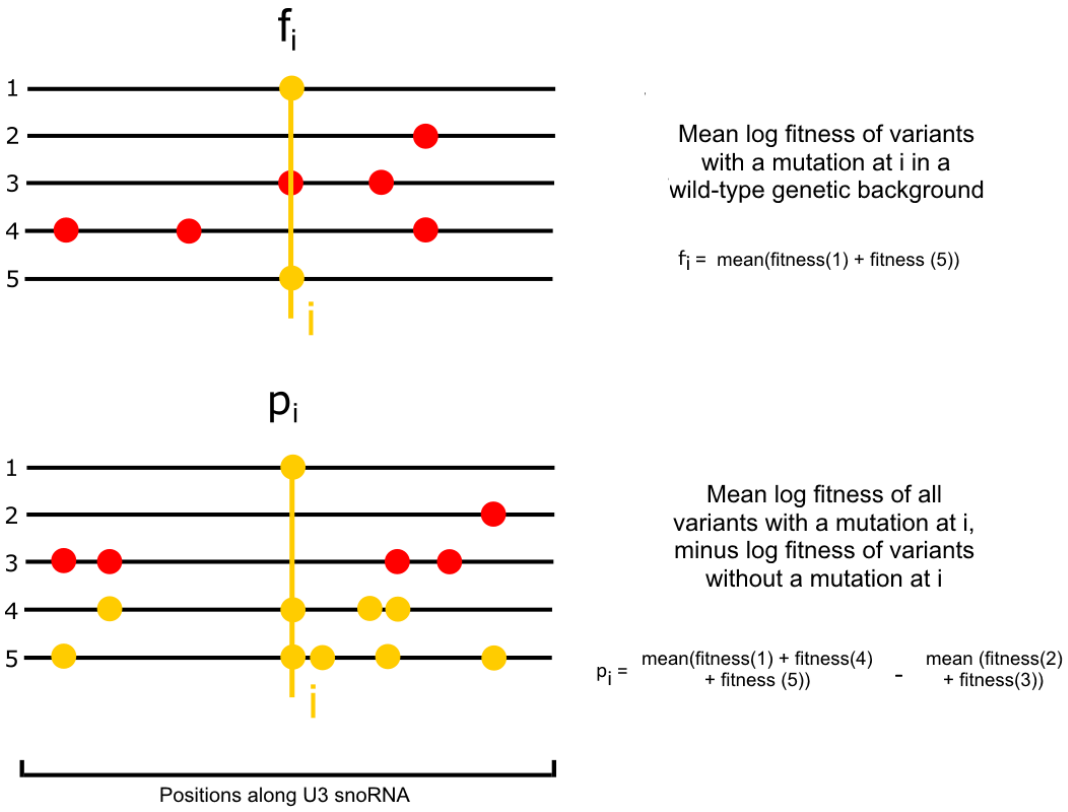


Figure 9, Calculating f_i and p_i values

To observe positional fitness effects f_i and p_i can be calculated for each position of U3 snoRNA. The f_i value for each position only takes into account mutations in a wild-type background, while p_i includes all genetic backgrounds. Five example genotypes are shown, with genotypes used in the calculation highlighted in yellow, and excluded genotypes in red.

With f_i the presence of single mutations was tolerated at most positions, reflecting mutational robustness (**Figure 10A**). The exception was the box B, box C, box D and box C' motifs, with mutations having a large detrimental fitness impact. These conserved sites allow the proteins that form the U3 snoRNP to bind, integral to U3 carrying out its role. Other results also matched functional expectations, with G-C base pairs in stems, thought to be important to RNA stability, being sensitive to mutation. A curious finding was a U->C mutation at position 191 that was highly deleterious, thought to create a consensus binding motif for Nab3 (UCUUG), a RNA degradation factor. This heterogeneity can be exposed due to the breadth of mutations covered in the library.

With p_i the there was an enrichment of negative fitness in more regions (**Figure 10A**). An example was the hinge region of U3 snoRNA, a region whose complementarity with the 5' ETS is used to direct pre-rRNA processing. This correlates with functional expectations, as base pairing requires multiple mismatches to ablate duplex formation. This pattern reflects the idea that fitness can change dramatically when a thermodynamic threshold is breached (Bershtein, et al., 2006). There were however positions that still tolerated mutations, such stem IV and VI, indicating these regions were not vital to U3's function.

When correlating f_i and p_i, the relationship was found to be non-linear (**Figure 10B**). This indicated many mutations had small effects in a wild type background, but much increased in the presence of additional mutations. It was also found that the average log fitness decreased as a function of increasing number of variant mutations, higher than would be expected if just aggregating the effect the single mutations. These patterns suggested that negative epistasis is abundant within the U3 snoRNA. When analysing the

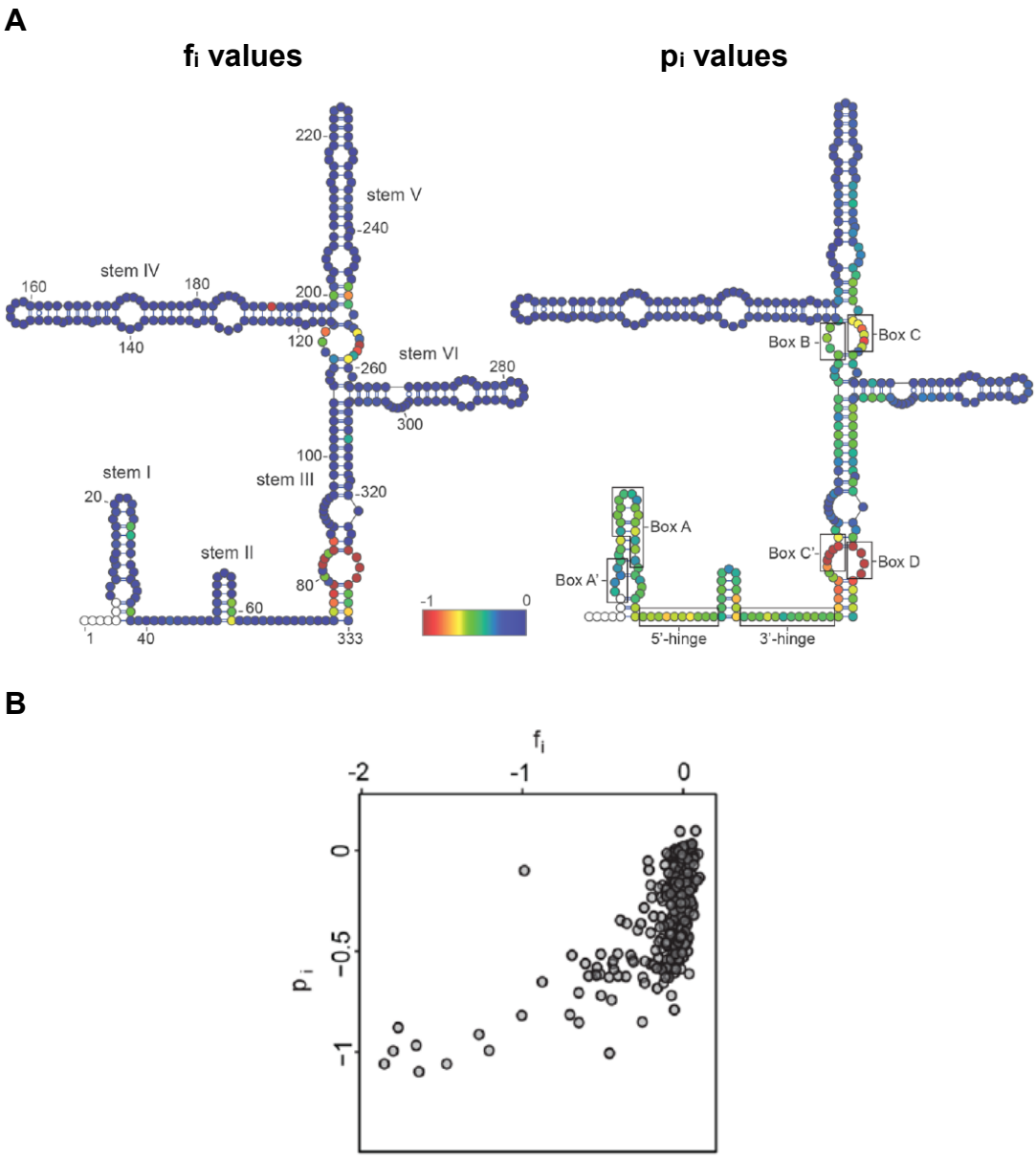


Figure 10, Competition experiment performed in physiological conditions

A. U3 snoRNA secondary structure with the log fitness effect of mutations (f_i and p_i) represented via a colour scale, with blue indicating no effect to red indicating a strong negative effect. Indicated also are nucleotide numbers, stem labels and black boxes highlighting characterised regions. (From (Puchta, et al., 2016)).
B. Correlation of f_i and p_i values, highlighting non-linear relationship. (From (Puchta, et al., 2016)).

network of epistatic interactions negative interactions were indeed found to be the most common, for example in the aforementioned hinge regions. There were however examples of positive interactions, such as compensatory mutations to restore base pairing, an example of reciprocal sign epistasis.

In this thesis I use this experimental system to elucidate genetic interactions within the fitness landscape of U3 snoRNA, helping understand the molecular mechanisms behind epistatic observations. In this chapter I will detail the modifications I made to enable the study of gene-environment and gene-gene interactions. This included increasing the coverage of the mutant library, optimising the efficiency of yeast transformation, effectively downregulating U3 snoRNA interacting proteins and altering the bioinformatic pipeline for processing the data.

3.2 Adding variants to U3 snoRNA mutant plasmid library

The U3 mutant library utilised in (Puchta, et al., 2016) had mutations encompassing 98% of the U3 snoRNA gene, from nucleotide positions 7-333. The library was created using doped oligonucleotides in an unbiased saturation mutagenesis approach, a key part of the deep mutational scanning aspect of the experiment, and is described in the Materials and Methods section of this thesis.

Nucleotide positions 1-6 of U3 snoRNA were not represented due to the nature of the library creation, however it is known that these positions are important in both pre-rRNA and protein interactions. This region contains the GAC box, which is essential for site 1 cleavage in *Xenopus* (Borovjagin & Gerbi, 2001). It was also found that mutations in this region were able to rescue a hypomorphic Dhr1 mutant, indicating that it is an *in vivo* target of Dhr1 (Sardana, et al., 2015). So in the interest of having a more complete U3 snoRNA mutant library, especially when studying genetic interactions with interacting proteins, I used site directed mutagenesis to add single mutations to this region.

3.2.1 Site-directed mutagenesis

To add the new variants to the library I mutagenized the pU3-wt plasmid. This differs from the vectors in the mutant plasmid library in that it contains the wild type U3 snoRNA sequence and does not contain a molecular barcode. I therefore had to both mutagenize the first six positions and insert a 20 random nucleotide barcode sequence at the designated position.

I initially attempted to use a whole plasmid site directed mutagenesis approach, using primers that were the reverse complement of each other. The six primer pairs contained a complimentary mismatch that would add a degenerate mutation to positions 1-6 of the U3 snoRNA sequence. These six PCR reactions would synthesise the complimentary strands of the plasmid with the desired position mutated. The Dpn1 restriction enzyme only digests methylated DNA, and so could then be used to remove the pU3-wt template while leaving the newly synthesised unmethlyated PCR products. The PCR products could then be ligated and cloned into competent DH5α *E. coli*. I found however that I obtained little PCR product and could not obtain colonies. This could be due large size of the plasmid (7825 base pairs) resulting in the PCR reactions never synthesising the full product, or not producing sufficient product to satisfying the requirements of the ligation reaction and subsequent transformation. This was still the case when repeated with Phusion, Accuprime and Q5 polymerases, and with different primer annealing locations.

An alternative approach was to isolate instead the region containing the U3 snoRNA sequence including where the barcode would be located, mutagenize, and then ligate back into the plasmid (**Figure 11A**). The smaller size of the fragment (885 base pairs vs 7825 base pairs) would make the PCR reactions more likely to succeed, and meant the mutagenesis and addition of a molecular barcode could occur in two parallel reactions rather than sequentially. The desired location of the barcode sequence in the plasmid was five nucleotides from an EcoR1 restriction site. I took advantage of this by designing a primer to anneal where the barcode would be, with the primer overhang containing the barcode and the sequence 18 nucleotides upstream including the EcoR1 site (Adding barcode F). The reverse primer annealed at positions 7-26 in the U3 snoRNA sequence on the plasmid (Adding barcode R). This 434 nucleotide PCR product would be used to add the molecular barcode. The second PCR to add the mutations used a primer

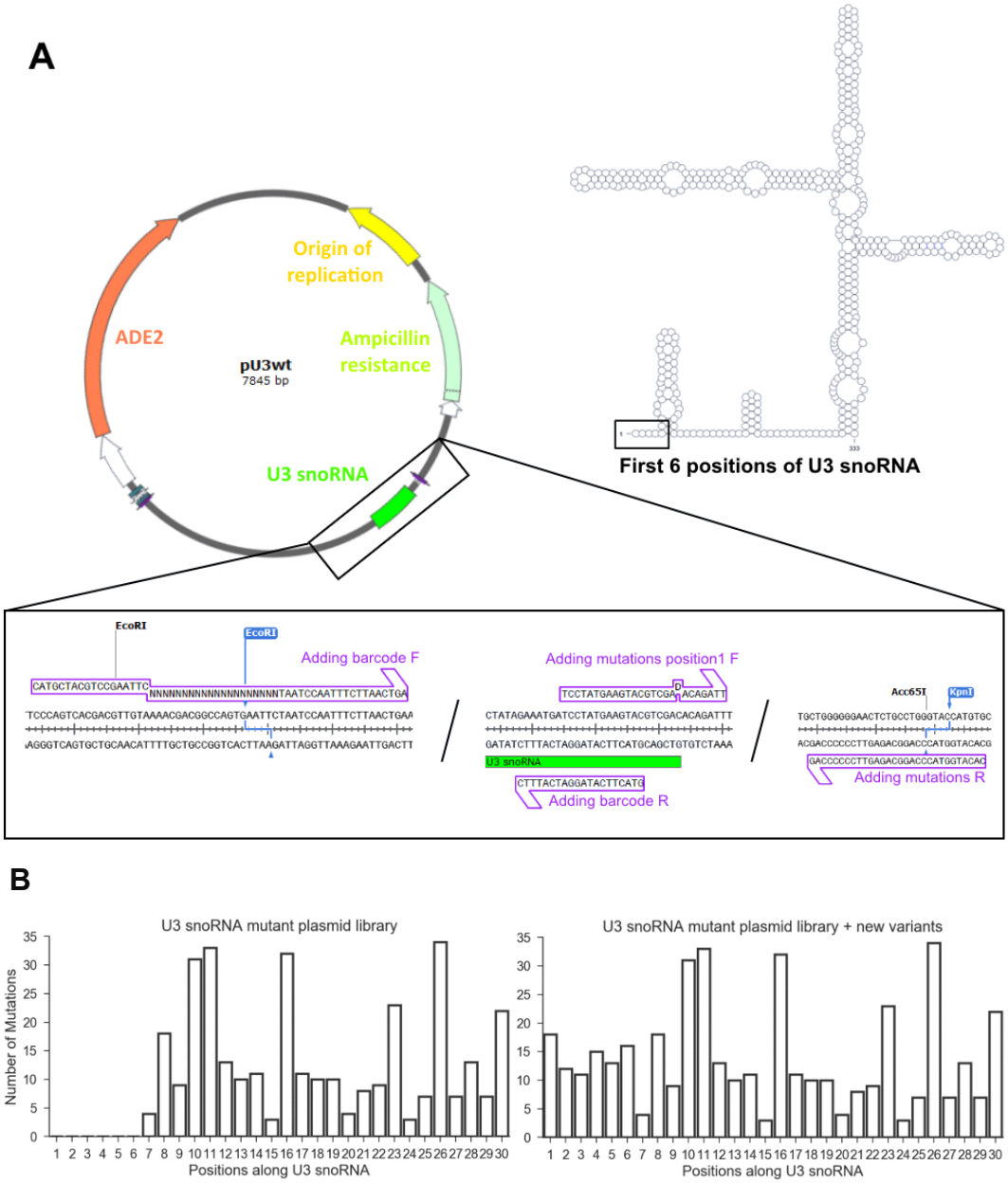


Figure 11, Adding mutations to U3 snoRNA mutant plasmid library

A. Schematic showing the secondary structure of U3 snoRNA, with the first six nucleotides highlighted with a black box. Also shown is a plasmid map of pU3-wt, with the adenine sequence (ADE2) in orange, origin of replication in yellow, ampicillin resistance cassette in mint and U3 snoRNA sequence in green. The location and strategy for site directed mutagenesis is expanded with a black box. The four primers are show in purple ("Adding mutations position 1 F" primer has a "B" at U3 snoRNA position 1), with the restriction sites indicated.
B. Histograms showing the distribution of single mutant variants in the first 30 positions of U3 snoRNA, in the original U3 "small" plasmid mutant library and after the addition of 80 new variants.

that annealed to the region covering the first six positions of the U3 snoRNA sequence, with six primers containing the single degenerate sequences for those positions (Adding mutations position 1-6 F). For example when ordering primers, using the IUPAC nucleotide code of ‘D’ at the first ‘C’ position of the U3 snoRNA complementary sequence would result in A, G or T. The reverse primer annealed further downstream, designed to encompass a Kpn1 restriction site, creating a 464 nucleotide product (Adding mutations R). These two PCR fragments shared 13 nucleotides of complementarity via positions 7-19 of the U3 snoRNA sequence, and so could be combined via an extension PCR reaction. The resulting 885 nucleotide oligo was digested with EcoR1 and Kpn1 and cloned into the accordingly digested pU3-wt plasmid in six reactions for the six positions. I recovered hundreds of bacterial clones for each position, with 16 of each position being Sanger sequenced in a 96 well plate. I obtained 80 successful sequencing reaction results, with every degenerate mutation represented multiple times for all six positions, and by chance three wild type sequences. All variants had also been assigned a random barcode sequence that did not already exist in the library.

I then had to combine the new mutant variants with the existing “small” U3 snoRNA mutant plasmid library. Originally I had planned to simply spike the plasmids in at a similar relative molarity to the existing plasmid library. Upon discussion with my lab however, I discovered that previously when spiking in wild type sequence plasmids isolated from a different experiment they were noticeably distinguishable when analysing the sequencing results. It is unknown what caused this, but may be due to the bacterial transformation during plasmid preparation occurring in a different strain of bacteria. To avoid this technical problem and maintain parity between all variants, I retransformed the existing “small” library in combination with the new plasmids into DH5α *E. coli*. This involved a large transformation, plating upon 100 plates with ~150,000 colonies. These were then pooled together in PBS, and the plasmids

purified with 8 MAXIprep columns. Since the 80 new plasmids added contained three plasmids with wild type U3 snoRNA sequence, I was able to use them as a control. I subsequently found these three variants performed similarly to the other ~3000 wild type variants in the library in every experiment, providing confidence for the other 77 variants added. The added variants bring positions 1-6 into parity with other single mutated positions in the library (**Figure 11B**), however do not provide the coverage of multiple mutations present for other positions in the library (**Figure 8C**). This means that while f_i can now be calculated for the first six positions, p_i still cannot.

3.3 DAmP strains

The ability to manipulate gene expression is a powerful molecular biology tool, and a cornerstone in the study of *in vivo* biological processes. This allows the assignment of gene function, elucidating cellular processes and pathways. While in many cases deletion or knockout of a gene provides the perturbation required, it is sometimes more appropriate to merely downregulate. It is also situationally advantageous to use a conditional system, to induce perturbation at specific times.

In yeast there are many methods to control gene expression, presenting a powerful eukaryotic model system. Temperature sensitive mutants have been used extensively, and contain a mutation that renders them unable to grow at a restrictive temperature. These however have the secondary effect associated with temperature change. Conditional promoters such as *GAL1*, *CUP1* or the Tet-OFF system provide powerful methods to manipulate gene expression, and do not require changing the gene sequence. However these systems require replacing the physiological promoter, show significant leakiness, and are limited by the turnover of mRNA, taking 12-24 hours to see complete depletion. Temperature sensitive degrons utilising the N-end rule allow fast, conditional degradation of the protein of interest, but again require a temperature shift to 37°C. Two methods that provide the ability to downregulate a protein of interest to hypomorphic levels are the DamP system and AID system, with the AID additionally being conditional.

3.3.1.1 Decreased abundance by mRNA perturbation

Investigating the mechanisms of intermolecular epistasis between U3 snoRNA and its interacting proteins required the disruption of said proteins. I set out to achieve this by downregulating the interacting proteins to hypomorphic levels, then carrying out the competition experiment under these conditions. The first method we tried to achieve downregulation was utilising a collection of decreased abundance by mRNA perturbation (DAmP) mutants. The DAmP library is a collection of mutants whose mRNA is compromised in such a way to downregulate the translated protein (Breslow, et al., 2008). These hypomorphic alleles are available for >950 essential yeast genes. The principle behind the method lies in the disruption of the 3' UTR by a kanamycin resistance cassette (Muhlrad & Parker, 1999). This leads to relative destabilisation of the transcript, reducing mRNA abundance 4-10 fold. It was also shown that this method is suitable at studying genetic interactions between two downregulated proteins (Breslow, et al., 2008). While not being conditional, this method does offer a way of studying essential genes in situations where a full knockout is too severe.

Yeast strains were purchased for U3 snoRNA interacting proteins Nop1, Nop58, Rrp9 and Imp4, along with cell wall maintenance protein Emw1 as a control. These strains could not be used in the competition experiment however as the competition experiment requires a strain in which U3 snoRNA is under the control a galactose inducible promoter. These strains were instead used as a template for PCR amplification. A region containing the kanamycin cassette and 500 nucleotides upstream and downstream were PCR amplified from DNA extracted from these strains. The 500 nucleotides either side acted as homology arms when transformed into the D343 strain, facilitating integration by homologous recombination at the gene of interest. This allowed the DAmP method of downregulation to be used in the D343 strain. This work

was performed prior to my involvement, with my first task being to characterise the phenotype of these strains.

It should be noted that unbeknownst to us, this D343 strain was in fact diploid, and thus the transformants were only heterozygous for the DamP insertions. As expected while heterozygous diploids may have reduced gene dosage, they often show no growth phenotype (Deutschbauer, et al., 2005). Subsequently the characterisations I performed were indistinguishable from the original D343 strain. Later on in this chapter I describe how this issue was resolved, obtaining haploid yeast.

3.3.2 Growth phenotype

The DAmP strains obtained from the DAmP collection for Nop1, Nop58, Rrp9 and Emw1 all reported a growth defect. I set out to assess if the D343 strain with DAmP alleles exhibited a similar phenotype by plating out dilutions of yeast and observing the colony size. While the original DAmP strains indeed showed a slight growth defect, the D343 strains did not. Their growth was indistinguishable from the original D343 strain (not shown).

3.3.3 Pre-rRNA levels

An established way to characterise a component of the pre-rRNA processing pathway is quantify different pre-rRNA species after depletion. Creating conditionally lethal alleles and observing the ratio of pre-rRNA species was used to implicate many proteins in the pre-rRNA processing pathway. In the case of proteins involved with U3 snoRNA, and thus early processing steps, it

was often observed that the 18S pathway via the first three cleavage events was impaired. The longer pre-rRNA species prior to these cleavage events would subsequently accumulate after depletion. Emw1 should show no processing defects when downregulated, serving as a control. In the absence of a growth defect, observing a processing defect was a prerequisite for completing the competition experiment to ensure U3 was affected and thus the DAmP alleles were working as intended.

3.3.3.1 RT-qPCR

The levels of different pre-rRNA species can be quantified using RT-qPCR. This involves creating cDNA using the total RNA from yeast as a template, then using the cDNA as a template for quantitative PCR. Primers were designed to amplify pre-rRNA species early in the processing pathway which should accumulate if U3's function was impaired. The '5' ETS' primer was designed to anneal between the start of the 35S transcript and the A0 cleavage site, detecting the full 35S transcript. The 'ITS' primer annealed between the A2 and A3 cleavage sites, detecting 23S, 27SA2 as well as the 35S transcript. The '5' ETS' and 'ITS' primer locations are indicated in **Figure 12**.The 'U3 snoRNA' primer was designed to anneal to the stem of U3 snoRNA, which should decrease if certain U3 RNP proteins are depleted. A primer for actin mRNA was used as an internal control. D343 grown in galactose represents an otherwise isogenic wild type control, while growth in glucose which strongly represses U3 snoRNA expression due to the GAL promoter, serves as a positive control. If the DAmP system was indeed reducing mRNA levels of the proteins of interest, I expected pre-rRNA levels to deviate from that of D343 in galactose, trending towards that of D343 in glucose. The strains were grown for six hours which is enough time for the onset of U3 snoRNA repression.

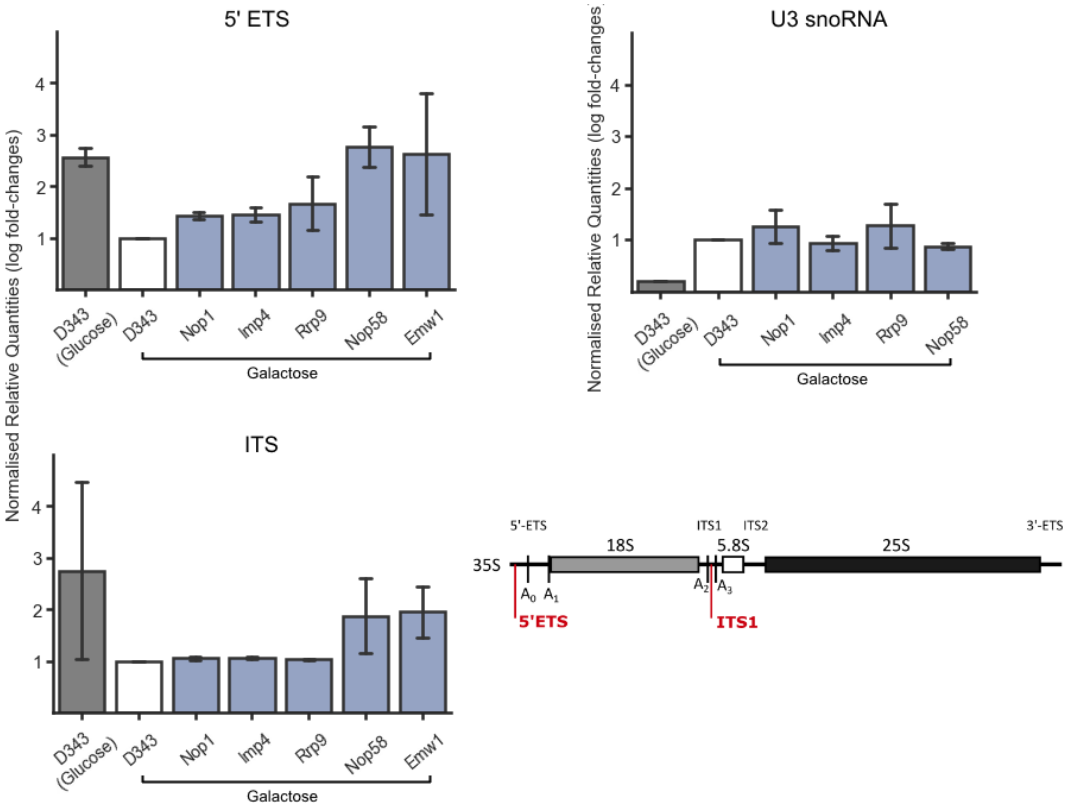


Figure 12, RT-qPCR characterisation for DAmP strains

Bar charts showing the log fold changes using primers in the ETS region or ITS region of the pre-rRNA, or for U3 snoRNA. The locations of the ETS and ITS primers on the 35S transcript are illustrated. Growth in glucose or galactose is indicated. The values were generated using the pfaffl method, with primer efficiencies calculated with a dilution series. The name of the gene indicates a strain with the DAmP allele.

Consistent with expectations, the U3 snoRNA steady-state levels are reduced when grown in glucose, with the ETS and ITS primers showing a resulting increased level of accumulated pre-rRNA (**Figure 12**). As expected retrospectively, due to the heterozygosity of the DAmP alleles there is no striking difference with these strains. Any deviations from the D343 in galactose sample seemed to be due to noise rather than a true difference. This is highlighted by any increase in the DAmP strains being matched by the Emw1 negative control. Nop58 which potentially shows accumulation of pre-rRNA, has no reduction in U3 snoRNA levels which would be expected (Lafontaine & Tollervey, 1999). This inconsistency was also observed when I attempted to use primers for the specific DAmP allele mRNAs, to directly quantify the reduction in mRNA levels (not shown). These results indicated that perhaps the DAmP allele phenotype was too subtle to detect, and in any case would be unsuitable to use in the competition experiment. Due to the noise however, I decided to quantify pre-rRNA levels without the multitude of processing steps associated with RT-qPCR using northern hybridization. This would more conclusively characterise the strains.

3.3.3.2 Northern hybridization analysis

Similar in concept to using RT-qPCR, Northern blotting can be used to directly visualise the RNA levels. I designed probes to hybridize in similar locations to the qPCR primers. The ‘5’ ETS’ probe hybridized between the start of the 35S transcript and the A0 cleavage site. The ‘ITS1’ probe hybridized between the 18S and A2 cleavage site. The ‘18S’ probe was used to show mature 18S levels. The ‘U3 snoRNA’ probe was used to show U3 snoRNA steady-state levels. The ‘SCR’ probe hybridized to SCR1, a RNA subunit of the signal recognition particle, and was used as a loading control. This method should allow direct visualisation of the different pre-rRNA species, facilitating a more

conclusive characterisation of the DAmP strains. When performing the RT-qPCR the phenotype even when U3 snoRNA expression was repressed was quite subtle. To account for this I grew the strains for eight hours rather than six, to produce a more pronounced phenotype for the D343 in glucose control. Two biological replicates were very similar, and thus I have just shown one in **Figure 13**.

As with the RT-qPCR the D343 control strain was grown in galactose and glucose, with the strong repression of U3 snoRNA expression by growth in glucose serving as a positive control. The northern hybridization analysis enables the ratio of different pre-rRNA species to be visualised. **Figure 13** shows that when U3 snoRNA expression is repressed via glucose the 35S transcript accumulates, shown by both '5' ETS' and 'ITS1' probes. Conversely the levels of the 20S species, which is processed to the mature 18S, decrease as the requisite cleavage events are inhibited. As expected, this same level of reduction is seen in mature 18S levels. The U3 snoRNA levels being reduced assures that the GAL promoter is indeed repressed. It should also be expected that $27SA_2$ levels should be decrease, however this species was not detected. The 23S species that accumulates is caused by the 35S transcript being cleaved at A_3, accumulating due to cleavage events A_0, A_1 and A_2 being inhibited. This also gives rise to the $27SA_3$ species that is not detected.

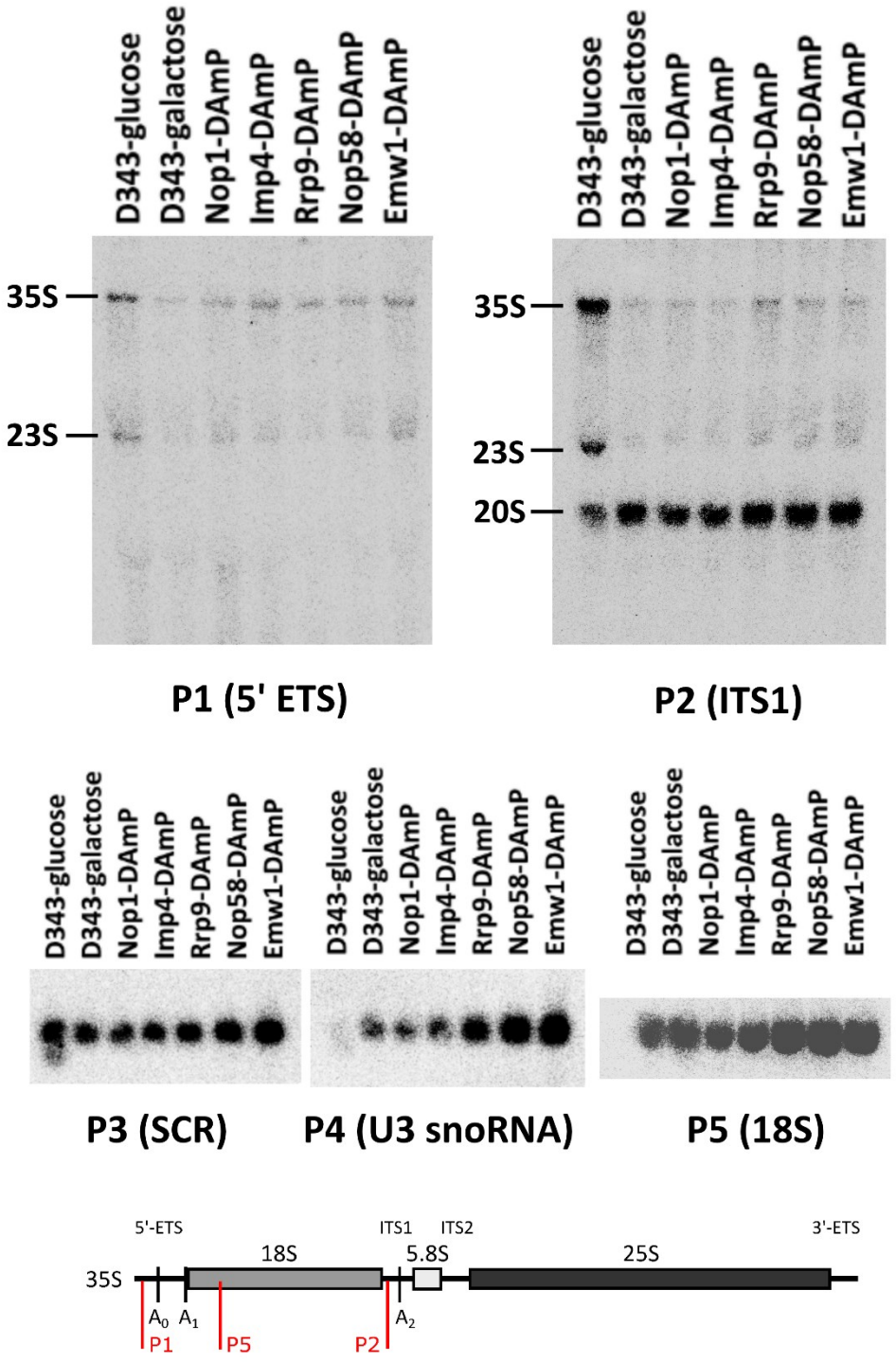


Figure 13, Northern hybridization in DAmP strains

Northern analysis of pre-rRNA, U3 snoRNA and SCR1 levels for D343 and DAmP strains. Probe 1 (P1) is for the 5'ETS of the pre-rRNA, Probe 2 (P2) is for the ITS1 region, Probe 3 (P3) is for SCR1, Probe 4 (P4) is for U3 snoRNA and Probe 5 (P5) is for the 18S region. The probe locations of P1, P2 and P5 are illustrated on the 35S transcript.

When observing the DAmP strains, all appear indistinguishable from the D343 galactose control. There appears to be no accumulation of aberrant species. There is a slight pattern of increasing quantities from the left to right side of the blot, however this reflected across all probes indicating slight nonuniformity in the amount of RNA loaded. I concluded from this result that the DAmP strains were either not reducing mRNA levels, or the reduction was too subtle to cause a sufficient phenotype. I therefore decided to try an alternative method of protein downregulation, with it appearing possible to use the auxin inducible degron system to reduce proteins to hypomorphic levels (Nishimura, et al., 2009).

3.4 Auxin inducible degron strains

The auxin inducible degron system provides a rapid, reversible and ajustable method of proteolytic elimination of a protein of interest. The system was adapted from plants, and has now been engineered to work in yeast (Nishimura, et al., 2009), *C. elegans* (Zhang, et al., 2015), *Drosophila* (Trost, et al., 2016) and mammalian cells (Natsume, et al., 2016) (Nishimura & Fukagawa, 2017). In its physiological context it regulates the transcriptional repressor IAA proteins to modulate growth. This is achieved by the auxin IAA facilitating an interaction between the degron domain of the IAA proteins and the F-box protein Tir1. After the interaction occurs within the recognition domain of Tir1 the IAA protein is ubiquitylated by SCF-type ubiquitin ligase (E3), and subsequently degraded. The SCF complex is highly conserved in eukaryotes, requiring only the addition of Tir1 to adapt the system to non-plant cells. Thus ectopic expression of Tir1 combined with a protein of interest containing a degron at the N or C terminus is all that is needed to establish auxin dependent depletion system (**Figure 14**). Degradation occurs within 20-30 minutes, with the level of depletion being dependant on auxin levels. The auxin used can either be the natural auxin IAA, or the synthetic NAA. These act very similarly, with NAA being more suitable to avoid growth defects in lower nutrient media (Nishimura & Kanemaki, 2014) or when using blue light for GFP excitation (Papagiannakis, et al., 2017). The downregulation can then be reversed with the removal of auxin. This system is relatively simple to execute in yeast, requiring only the transformation of PCR amplified DNA (Nishimura & Kanemaki, 2014). This system aims to minimise the secondary effects, with Tir1 expression being biologically silent. The main limitation of the system is certain proteins displaying low levels of degradation, with it currently being unknown what causes this variation in phenotype.

The system was first improved by identifying the minimal amino acids of IAA17 (229 aa) necessary for the degron to work with either the mini-AID (65-132 aa) (Kubota, et al., 2013) or AID* (71-114 aa) (Morawska & Ulrich, 2013). The degron was also augmented with epitope tags for detection by fluorescence microscopy or antibodies (Morawska & Ulrich, 2013 , Tanaka, et al., 2015). To tackle the problem of incomplete depletion and create tight knockout mutants, the Tet-OFF system was incorporated and termed the improved auxin-inducible degron (iAID) (Tanaka, et al., 2015). This combination of transcriptional repression and protein depletion was found to help even with the most difficult proteins.

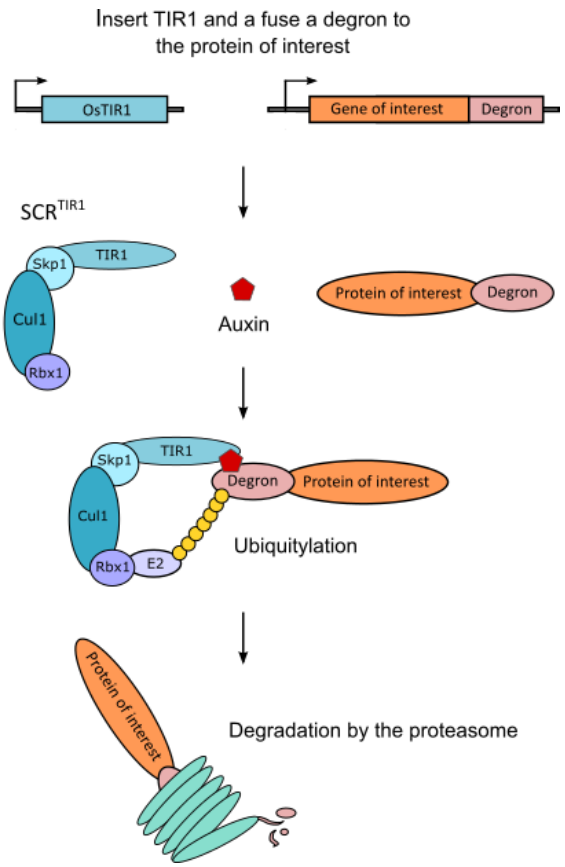


Figure 14, Auxin inducible degron system

The auxin inducible degron system requires expression of TIR1 and a degron tag on the protein of interest. In nonpermissive conditions (presence of auxin) the tagged protein is ubiquitylated and rapidly degraded.

This allowed strategies such as adding both auxin and doxycycline to the media, or even pre-depleting proteins with low levels of doxycycline. Other methods to deal with the issue have been to use a mutation to destabilise a protein to facilitate more efficient degradation by the auxin system (Watase, et al., 2012). Strategies have also included regulating TIR1 via a β-estradiol expression system, with conditional expression allowing a more tightly controlled system (Mendoza-Ochoa, et al., 2019). The system was recently extensively quantified in single yeast cells, allowing detailed characterisation (Papagiannakis, et al., 2017). This highlighted that certain auxin levels may be sufficient for depletion, but required concentrations >250mM to minimise intercellular variation. It also showed that after removing auxin, post depletion takes multiple generations to reach the physiological equilibrium.

The creation of these mutants is made more straight forward by the use of a PCR amplification protocol. This methodology of transforming yeast with PCR amplified DNA to create AID mutants is described in (Nishimura & Kanemaki, 2014), and in the Materials and Methods section of this thesis. The basis of the protocol is PCR amplifying the AID sequence from a plasmid using primers with a 50bp overhang homologous to the gene of interest. This can then be transformed into yeast, and homologously recombined at either the C or N terminus of the gene. This can also include the insertion an inducible promoter.

It is possible to purchase strains already expressing Tir1, however for use in the competition experiment U3 snoRNA must be under a galactose inducible promoter. In order utilise the AID system for the competition experiment, I first inserted Tir1 at the Can1 locus, subsequently inserting AIDs after the genes of interest. The most frequent uses of the AID system take advantage

of it being both rapid and conditional, offering a unique tool in this respect. For my requirements however, the ajustable nature of the system is most useful, as it has been shown that protein depletion can change as a function of auxin levels (Nishimura 2009). In this way I planned to use the AID system to downregulate the proteins of interest to hypomorphic levels.

3.4.1 Transforming TIR1

The first stage in creation of the AID strain requires the heterologous expression of Tir1, the only component of SCF E3 ubiquitin ligase that is not conserved in yeast. I used the pMK200 plasmid which contained the sequence of Tir1 from *Oryza sativa* (OsTIR1), which was found to be more functional and work at a more diverse range of temperatures than Tir1 from *Arabidopsis thaliana* or *Gossypium hirsutum* (Nishimura, et al., 2009). This sequence of Tir1 is codon optimised for yeast expression, and is controlled by the constitutive ADH1 promoter. I chose to insert it at the Can1 locus, a plasma membrane arginine permease which is mutated in the D343 strain. This would allow expression of Tir1 in the D343 strain with no secondary effects.

pMK200 contained no resistance cassette, and instead was an integrative vector, designed to be linearized with the StuI restriction enzyme digestion and inserted at the URA3 locus via homology. This could not be performed in the D343 strain however, as URA3 was the location of U3 snoRNA under control of a galactose-inducible promoter. So instead I PCR amplified a hygromycin cassette from pSM409, and inserted it into the pMK200 plasmid. To do this I first used primers to PCR amplify the region containing the hygromycin cassette and a Sac1 restriction site, with one primer having an overhang containing a Not1 site. The resulting PCR fragment could be digested with

Not1 and Sac1 and cloned into the pMK200 plasmid, neighbouring the OsTir1 sequence. This provided a template containing both the OsTir1 sequence and a hygromycin resistance cassette. I PCR amplified a fragment containing these two sequences, with the primers containing 50 nucleotide overhangs with homology to the Can1 locus. After transforming into yeast and selecting on hygromycin plates, I used colony PCR genotyping to confirm insertion of Tir1 at the Can1 locus, PCR amplifying a fragment across the insertion site. The band sizes indicated both a wildtype and insertion allele due to the strain being diploid (**Figure 15A, lane 2**). This was my first indication that the strain was indeed diploid, however I decided to proceed with transforming the degrons for the proteins of interest. I grew both the D343 and D343-Tir1 strain on auxin plates to confirm no Tir1 dependent growth defects.

3.4.2 Transforming auxin inducible degrons

After obtaining a strain with heterologous expression of the Tir1, I next C-terminally tagged my proteins of interest. This involved PCR amplifying the 1xmini AID and kanamycin resistance cassette from the pMK68 plasmid, with each primer pair having overhangs with homology to the gene of interest. The homology locations were designed to remove the genomic stop codon, and place the degron within the open reading frame. I had plasmids containing both a 1x and 3x mini AID, with the 3 copies of the degron making up the 3x mini AID resulting in a stronger protein depletion effect. I decided to first try the 1x mini AID as I was aiming for hypomorphic levels and ideally wanted a small tag. The proteins chosen were U3 snoRNP proteins (Nop1, Nop56, Nop58, Snu13, Rrp9), Mpp10 complex proteins (Imp3, Imp4, Mpp10) and the helicase Dhr1. These proteins were chosen due to their direct interaction with U3 snoRNA being previously characterised. I also chose the

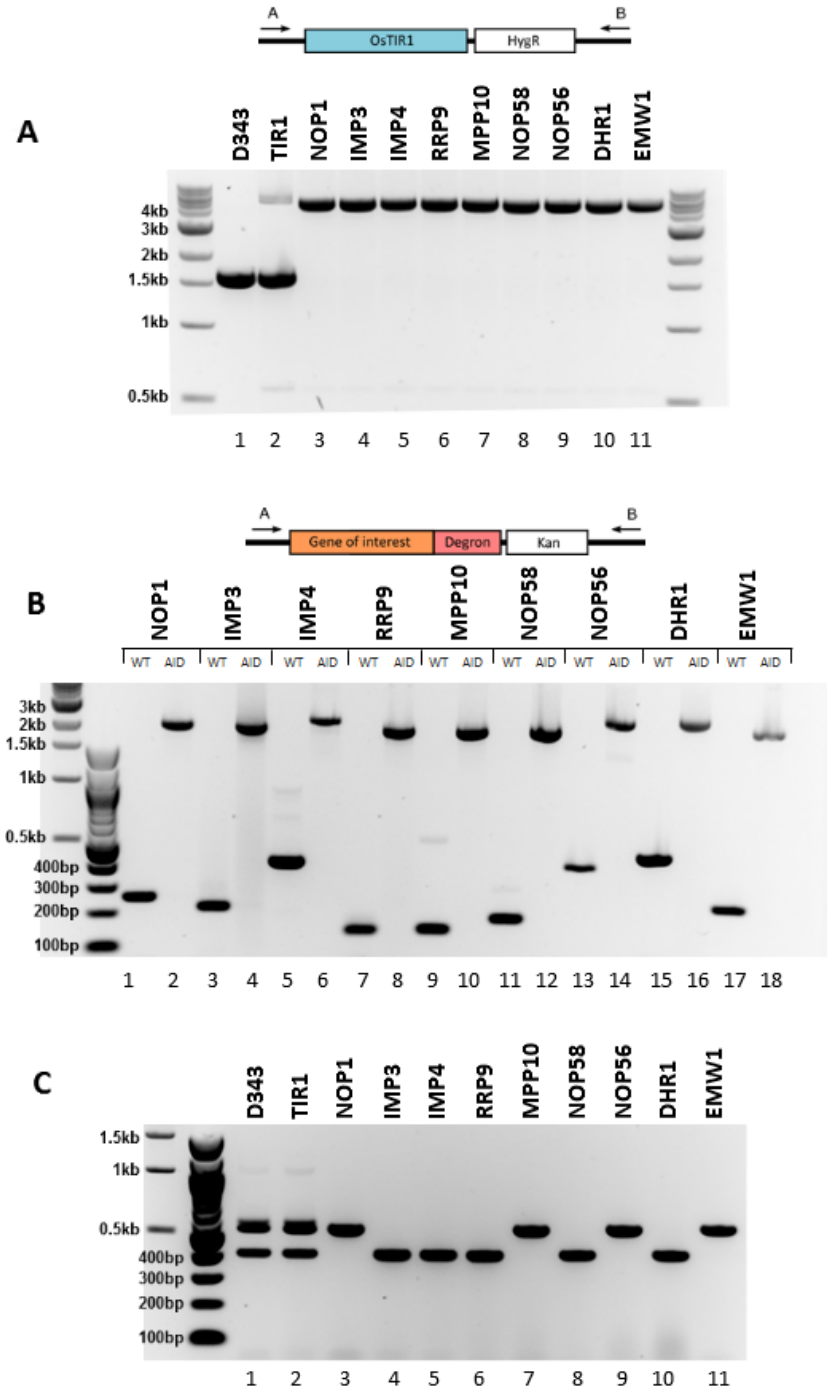


Figure 15 Genotyping PCR for AID strains (1% agarose gel)

D343 strain, D343+Tir1 (TIR1), and D343+Tir1 strains with degron at gene labelled.

A. PCR amplification across the insertion site of OsTIR1 in the Can1 locus.

B. PCR amplification across the degron insertion site at the 5' of genes of interest, with both the D343 strain (WT) and the AID strain.

C. PCR amplification at the MAT locus designed to reveal the mating type of yeast.

cell wall maintenance protein Emw1 to serve as a control protein, uninvolved in ribosome biogenesis.

After transforming into yeast, and selecting on kanamycin plates, I used colony PCR genotyping to confirm insertion of the AID at the genes of interest. All ten strains appeared to have both wild type and insertion alleles (not shown), due to diplody. This was also demonstrated by performing a three primer PCR designed to amplify products at the MAT locus, indicating the mating type. These strains had two bands, indicating they were both 'a' (404 bp) and 'α' (544bp) mating types (diploid). This pattern is illustrated in the MAT locus PCR is for D343 and D343-Tir1 diploid strains in **Figure 15C, lane 1+2**. Consequently there was no auxin dependent impact on growth, with the heterozygous AID strains growing the same as the wild type D343 strain on auxin plates (not shown). The next step was then to derive haploids from these diploid strains.

3.4.3 Sporulation and isolation of haploids

The PCR amplification products across the insertion locus indicating heterozygosity, and the MAT locus PCR showing indicating both mating types, indicated the strain was diploid. This was further proved by no AID strains showing auxin dependent growth defects, despite containing Tir1 and an AID after proteins of interest. It is fortunately relatively simple to obtain haploids by inducing sporulation in yeast, growing them in a non-fermentable carbon source with nitrogen starvation (Neiman, 2011). This causes the yeast to undergo meiosis, splitting into four haploid nuclei called ascospores within the ascus. A haploid cell can then be separated, cultivated in normal nutrient conditions, with the subsequent colonies being comprised of haploid yeast.

After sporulation I used colony PCR genotyping to confirm haploidy. Sporulation appeared to have worked as there was only one allele of OsTir1 at the Can1 locus (**Figure 15A, lanes 3-11**) and one allele of the AID at the nine genes of interest (**Figure 15B, lanes 1-18**). The MAT PCR also showed one mating type (**Figure 15C, lanes 3-11**). Interestingly by random chance the haploid strains I obtained had a range of mating types. This could be utilised in the future, as by mating two strains of opposite mating type the resulting diploid would have degrons on two proteins. This for example would be interesting to downregulate both Nop56 and Nop58 as they work together as a heterodimer (Cheng, et al., 2017). I also created a haploid strain containing just Tir1 at the Can1 locus, so if any future strains needed to be created it would require only the transformation of an AID into this haploid Tir1 strain. I was unable to obtain a haploid AID strain for Snu13, with there being only a small amount of diploid colonies after sporulation. The seeming inability for yeast to be viable with the functional AID system indicates Snu13 disruption cannot be tolerated, likely due to its essential role in multiple processes. This may stem from leakiness in the system, causing degradation in the absence of exogenous auxin, or just from the physical presence of the AID disrupting Snu13's function. This could be remedied by putting the AID at the N-terminus, or creating a DAmP strain.

I also used Sanger sequencing to confirm the sequence of the inserted AID and the surrounding locus. All nine strains had no mutations, with the genomic stop codon removed, replaced with the AID sequence and a new stop codon.

3.4.4 Growth phenotype

After genotyping that the strains were correct, I next tested their growth at different concentrations of auxin. I was using the synthetic auxin NAA rather

than IAA, as it had been reported that IAA may cause a growth defect on synthetic complete media (Nishimura & Kanemaki, 2014). The concentration of auxin used depends greatly on the protein of interest and the desired depletion. There has been a large amount of variability reported when using the AID system, with some proteins being resistant to proteolysis even at high levels of auxin (1mM), and others only requiring 75 μM for compete degradation. It was found by looking at single yeast cells that at least 250 μM was required to reduce intercellular variation (Papagiannakis, et al., 2017).

With the AID strains I have created, low (~100μM) levels of auxin were not sufficient for a noticeable growth defect, perhaps due to protein's location in the nucleolus. Through testing various concentrations I chose 350μM as I found it was the best compromise, causing a minimal degron independent growth defect but noticeable degron dependent growth defects. **Figure 16** shows serial dilutions of yeast in 0μM, 350μM and 700μM concentrations of auxin. At 350μM there is a clear degron dependent growth defect, while at 700μM a significant degron independent growth defect can be seen. While rich media can alleviate this degron independent growth defect, only around 20% nonessential gene deletions cause a measurable growth defect in rich media (Giaever, et al., 2002). The competition experiment is therefore carried out in minimal media to increase sensitivity, allowing for more the detection of subtle growth phenotypes. If a full knockout was desired with these strains, using a richer media to alleviate the degron independent growth defect may be required. It is also noticeable that there are auxin independent growth defects in the AID strains. This is most likely either due to leakiness in the system, disruption caused by the tag, or low levels of endogenous auxin. For my purpose this is not an issue, as I require constitutive not transient downregulation. It is also interesting that Mpp10 has a significant growth defect independent of auxin, suggesting that the tag itself is disrupting Mpp10's function. The structured C terminus of Mpp10 is involved in stabilisation and

bridging within in the SSU processome and may not tolerate a degron at this terminus. This could be perhaps be solved by placing the degron at the N terminus, or creating a DAmP strain.

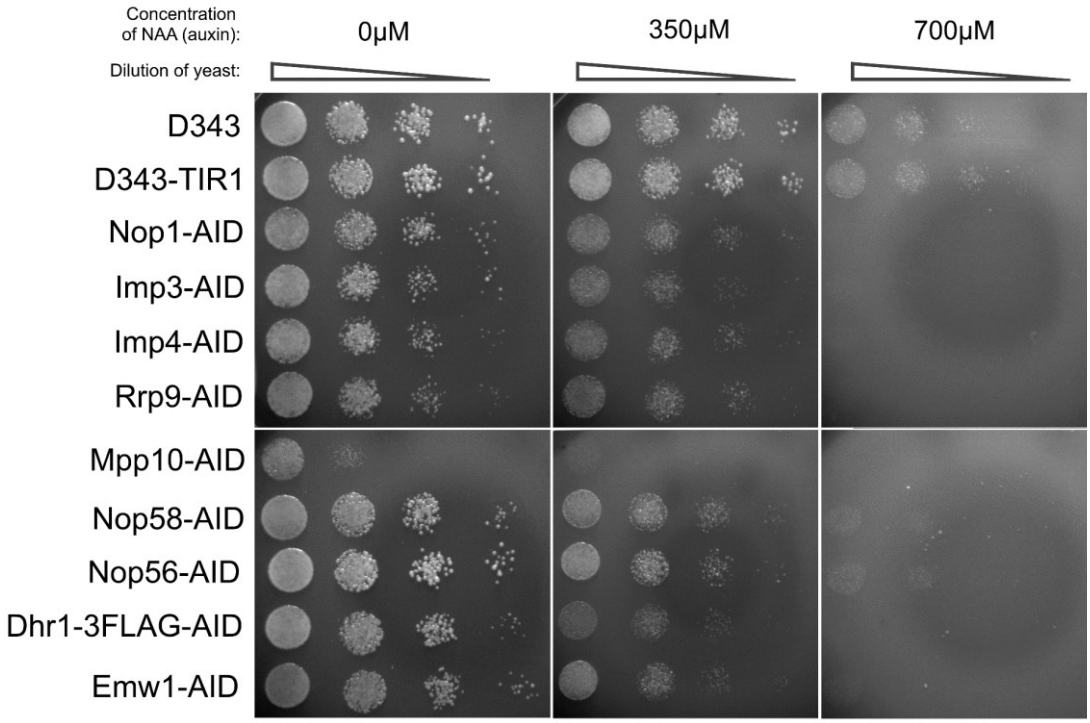


Figure 16, Dilution spot test of AID strains on NAA plates

Yeast plated out in 10μl spots, beginning at an OD_{600} of 0.1 and being diluted ten fold four times, and grown at 30°C for 3 days. The plates are CSM plus the indicated concentration of NAA.

3.4.5 Western blotting

The most direct way to observe downregulation is to directly measure the protein level. Western blotting involves the detection of protein levels using protein specific antibodies, facilitating semi-quantitate estimations by the size and intensity of protein bands. When first creating the AID strains I attempted western blots using an antibody for the degron tag itself. While I was able to see clear bands for the α-tubulin loading control, I could never observe bands for the degron tag in the Nop1-AID, Imp4-AID or Rrp9-AID strains. When initially transforming the AIDs into the Tir1 strain I obtained positive colonies for all proteins except Dhr1. I thus decided to transform both an AID and 3FLAG tag when retrying to create the Dhr1 AID strain, a strategy already shown to complement the AID system (Morawska & Ulrich, 2013). This would allow western blotting using an antibody for the FLAG tag, a well established epitope for western blotting.

3.4.5.1 3FLAG tag

When transforming the AIDs, I PCR amplified the degron sequence from the pMK68 plasmid, with the primer overhangs containing homology for the gene of interest. To incorporate a 3FLAG tag, I used site directed mutagenesis to add the 3FLAG sequence into the open reading frame of the degron sequence on the pMK68 plasmid. I utilised a similar strategy I initially used when first attempting to add mutations to the U3 snoRNA plasmid mutant library. I used two reverse complement primers annealing at the 5’ end of the degron sequence, whose two overhangs constituted two halves of the 3FLAG sequence. The 5’ end of these primers were phosphorylated, facilitating the PCR product to be ligated and circularised, resulting in the incorporation of the 3FLAG sequence. This ligation was then cloned with DH5α competent bacteria. I believe this strategy worked for pMK68 as it is 4176 base pairs, compared to

the 7845 base pair pU3-wt plasmid. This plasmid was then used to PCR amplify the AID and 3FLAG sequence to create the Dhr1-AID strain.

The insertion was checked by Sanger sequencing, with Dhr1 showing no mutations, and the 3FLAG tag being in frame with the degron.

I then performed western blots using an antibody against the FLAG tag. **Figure 17** shows the quantified levels of Dhr1 in different auxin conditions, with the lighter of the two bands being Dhr1. I cultured Dhr1 for one hour in 0 μM, 50 μM, 100 μM, 350 μM, and in 350 μM for 0, 15, 30 and 60 minutes. As expected increasing the concentration of auxin leads to increased protein proteolysis. At 350 μM it appears to cause a 70-80% reduction in steady state levels for Dhr1, which results in the growth defect observed. Also as expected the speed at which the protein is degraded is rapid, causing ~50% reduction after just 15 minutes. The maximum depletion for this concentration appears to have been reached by 30 minutes.

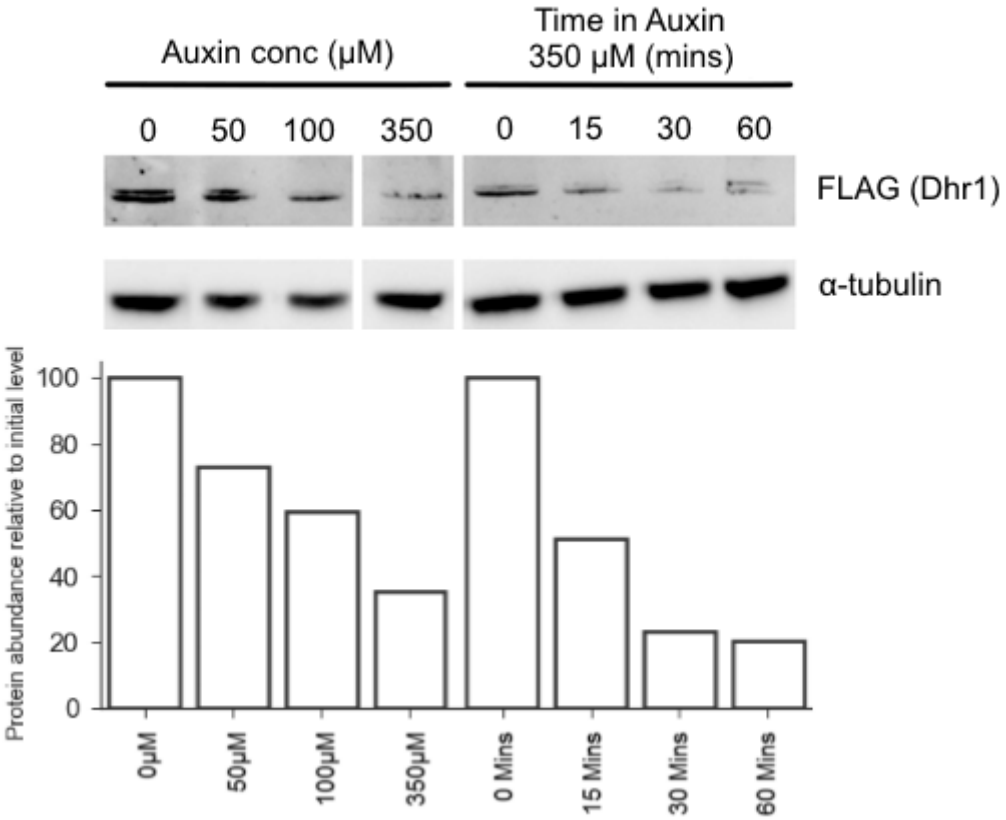


Figure 17, Quantifying protein proteolysis by western blotting

Anti-FLAG and anti α-tubulin antibodies, with the Dhr1-AID strains grown at different concentrations (0 μM, 50 μM, 100 μM, 350 μM) and for different periods of time (0, 15, 30 and 60 minutes). The Dhr1 band intensities (lower band) are quantified in a bar chart below, relative to either 0 μM or 0 minutes, and normalised to both the background and α-tubulin.

3.4.6 Pre-rRNA levels

Using the same methodology as with the DAmP strains, I quantified the levels of pre-rRNA species to assess the affect upon the pre-rRNA processing pathway. **Figure 18** shows the results of the RT-qPCR for Nop1-AID, Imp4-AID and Rrp9-AID strains and D343 controls. The '5' ETS' primer anneals before the A0 cleavage site, detecting the full 35S transcript. The 'ITS' primer annealed between the A2 and A3 cleavage sites, detecting 23S as well as the 35S transcript. The AID strains also show an intermediate accumulation of pre-rRNA, not as pronounced as when U3 snoRNA expression is repressed by glucose. This is as expected, as the strains are able to grow and thus U3's function is not completed ablated. There also appears to be auxin-independent changes in the AID strains, which reflects the growth defect observed in permissive conditions. Auxin appears to have little effect on pre-rRNA levels in the D343 strain, indicating it is biologically silent. Nop1 shows a low level of accumulation. This may be due to eight hours not being enough time for the phenotype to manifest, or there being increased robustness to the levels of Nop1.

In order to fully characterise these AID strains, as well as the others I have created, I intended to repeat this RT-qPCR after growth in auxin for 24 hours. This would also better reflect the state during the competition experiment, and expose the consequence on pre-rRNA processing. I also intended to use the same Northern hybridization analysis I used for the DAmP stains. Unfortunately due to time constraints this was not possible. Considering the genotyping (**Figure 15**), auxin dependent growth defects (**Figure 16**), the qPCRs completed (**Figure 18**) and the western blotting (**Figure 17**) I felt confident enough to use these AID strains in the competition experiment.

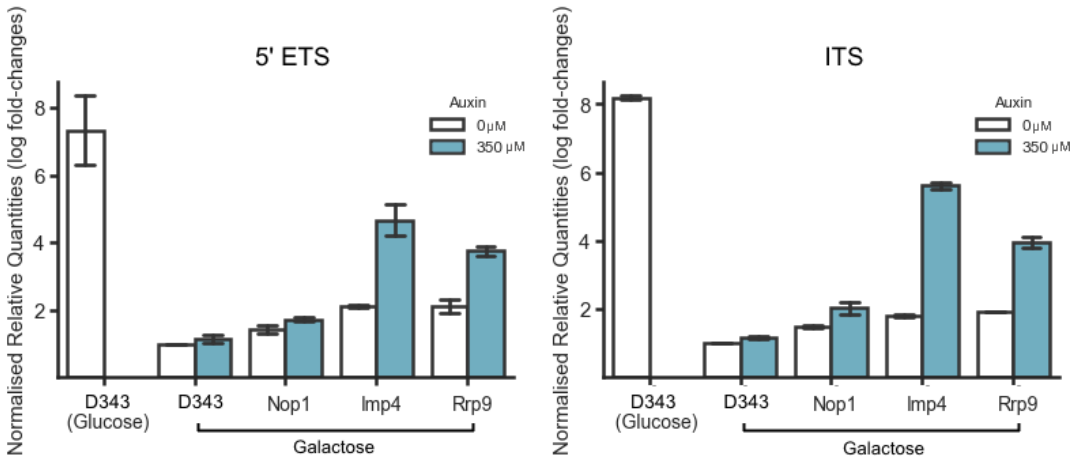


Figure 18, RT-qPCR characterisation for AID strains

Bar charts showing the log fold changes using primers in the ETS region or ITS region of the pre-rRNA. White bars indicate growth in normal conditions, teal indicating growth in 350 μM of auxin. Growth in glucose or galactose is indicated. The locations of the ETS and ITS primers on the 35S transcript are illustrated in **Figure 8**. The values were generated using the pfaffl method, with primer efficiencies calculated with a dilution series. The name of the gene indicates a strain with the AID allele.

3.5 Yeast transformation efficiency

An important aspect of the competition experiment is the transformation efficiency of the plasmid mutant library. It is not sufficient to simply equate the number of transformants to library size (~40,000) since the copy number of variants in the library are unequally distributed. In (Puchta, et al., 2016) the transformation efficiency achieved was ~6250 colonies per µg of DNA, which is much lower than the 1 x 10^6 per µg of DNA reported in (Gietz & Schiestl, 2007). This indicated the D343 strain was not a good transforming strain, requiring a large scale transformation to reach ~500,000 transformants. The initial competition experiments I completed (at 18°C with the D343 strain, and with the Imp4-AID and Rrp9-AID strains) did not achieve this (160,000, 145,000, 210,000 transformants respectively), but as it still seemed the majority of variants would be represented I proceeded to sequencing.

When processing the sequencing data ~3000 wild type variants within the library are used to estimate noise as function of coverage. They are utilised to set the minimal number of reads at the initial timepoint needed to assure reliable fitness estimates. This read number is then filter being applied to all variants, and the median wild type log fitness value is then subtracted from the other variants, normalising neutral fitness estimates to 0.

The distribution of wild type variants in the competition experiments I completed are shown in **Figure 19**, plotted to visualise the relationship between variation and the number of reads at the first timepoint. This can be thought to represent the inherent noise for a given experiment, as all wild-type variants should have the same log fitness effect. The variation decreases as a function of read count number, highlighting the necessity of filtering by this variable. It also visualises the variation of the 18C, Imp4-AID and Rrp9-AID

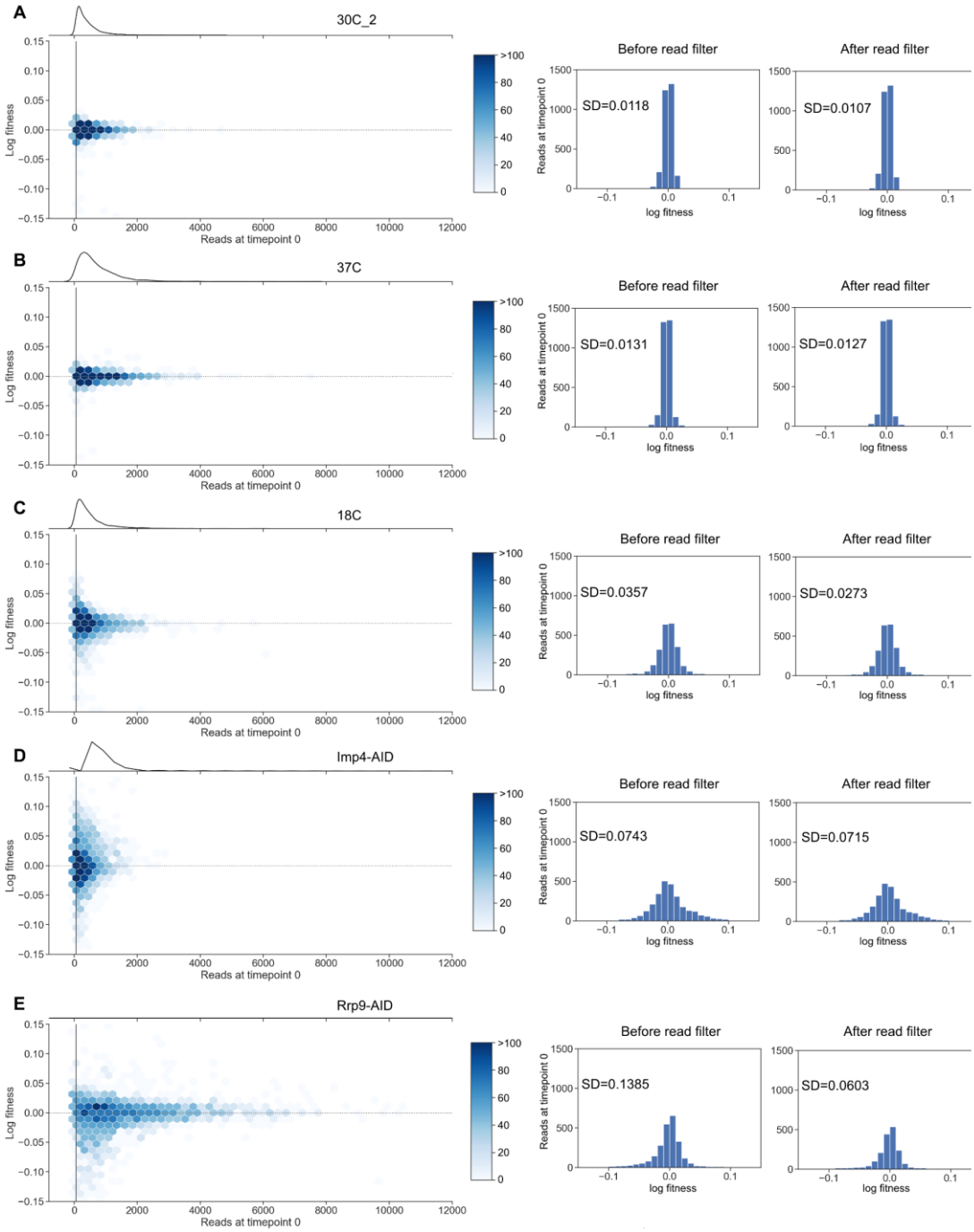


Figure 19, Noise as a function of coverage using wildtype variants, before optimisation

Hexbin density plots for 30C_2 (**A**), 37C (**B**), 18C (**C**), Imp4-AID (**D**), Rrp9-AID (**E**), visualising noise as function of coverage, with log fitness on the y axis and reads found at the first time point for each barcode on the x axis. A line indicates the read count filter, and a darker shade of blue indicates more barcodes per hexagon. The same data is also plotted in a distribution plot for each experiment before and after read count filtering with the standard deviation value. The values shown have been normalised to 0 by subtracting the median log fitness value obtained after read count filtering.

experiments relative to the other experiments.

As shown in **Table 1** the majority of variants were still detected in the 18C, Imp4-AID and Rrp9-AID experiments, only being 5-16% less than the “Small_3_30C_Glu” (30C_2) experiment from (Puchta, et al., 2016). The standard deviation between wild type variants before filtering however is much higher for Imp4-AID (0.0743) and Rrp9-AID (0.1385). It appeared that perhaps a higher number of transformants could help decrease this variation. This can be due to high transformation efficiencies allowing each variant to be transformed into more yeast, reducing intercellular variability associated with the age of a yeast cell (Jazwinski, 1990). The 18C and Imp4-AID experiments were sequenced by the HiSeq platform (1x50bp, ~200 million reads), the same as in (Puchta, et al., 2016). The Rrp9 experiment had higher read counts due to sequencing being completed by the NextSeq platform (1x75bp, ~400 million reads), but as detailed in (Matuszewski, et al., 2016), higher read counts are saturating beyond a certain coverage. (**Figure 19E**).

Therefore before carrying out further experiments, I set out to first improve the transformation efficiency to >6250 transformants per µg of DNA and increase the number of transformants to at least 500,000.

Table 1, Transformation statistics and log fitness variation

Experiment	Number of transformants	Number of reads used	Number of barcodes recognised	Number of barcodes after filtering	Standard deviation before filtering (wild type variants)	Standard deviation after filtering (wild type variants)
Experimental data from (Puchta, et al., 2016)						
30C_1	~500,000	76,505,765	41,617	19,367	0.0416	0.0328
30C_2	~500,000	165,773,130	41,359	24,312	0.0118	0.0107
37C	~500,000	88,384,609	41,190	23,108	0.0131	0.0127
Experiments using unoptimised transformation						
18C	145,000	86,493,258	34,505	19,565	0.0357	0.0273
Imp4-AID	165,000	129,967,084	38,395	20,938	0.0743	0.0715
Rrp9-AID	185,000	311,869,736	39,144	15,075	0.1385	0.0603
Experiments using optimised transformation						
D343_auxin	2,820,000	303,922,346	40,986	24,370	0.0284	0.0164
Nop1-AID	3,620,000	256,887,298	40,572	24,755	0.0314	0.0295
Nop58-AID	3,100,000	297,187,827	40,708	24,709	0.0226	0.0224
Nop56-AID	6,130,000	316,369,728	41,051	24,748	0.0200	0.0197

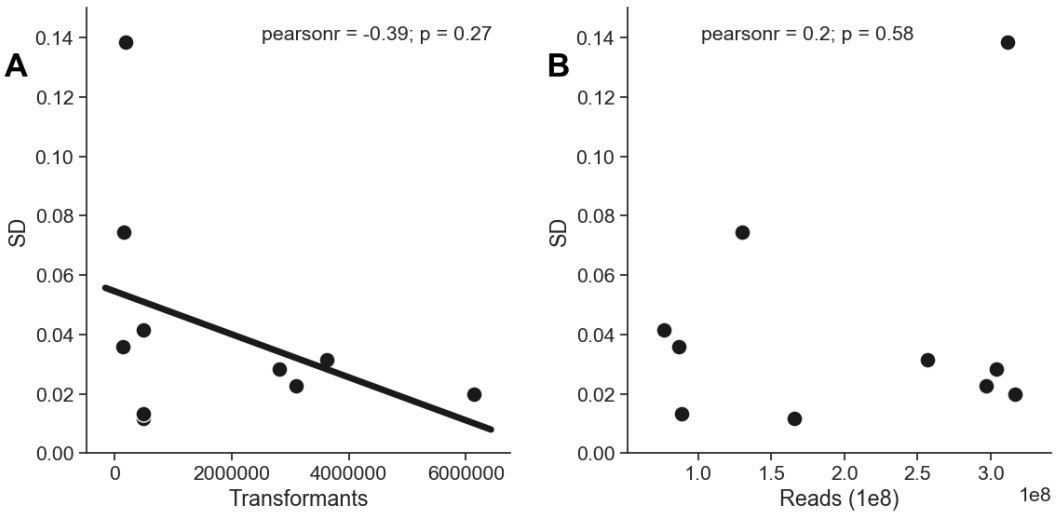


Figure 20, Correlation with the standard deviation of wild type variants

Correlation of the standard deviation of the wild type variants before filtering, with number of transformants (**A**) and number of reads used (**B**).

3.5.1 Transformation optimisation

The competition experiment transformation is effectively 30 fold larger than a single yeast transformation described in (Gietz & Schiestl, 2007). I performed my optimisations on x1 scale, and extrapolated the results to estimate how they would manifest for the competition experiment scale transformation. The five conditions I chose were:

- 2-fold less plasmid DNA
- 2-fold more plasmid DNA
- Incubating at 30°C for 30 min before heatshock,
- Resuspending and incubating for 20 minutes in LiAc (100 mM) before adding the transformation mix
- Resuspending in salmon sperm and plasmid DNA before adding the rest of the transformation mix.

The two different concentrations of plasmid DNA will highlight if this is a limiting factor. Incubating at 30°C and resuspending in lithium acetate are both thought to increase the permeability of the yeast cells (Kawai, et al., 2010) . Resuspending first with the single stranded salmon sperm and plasmid DNA before adding the rest of the transformation mix is thought to allow the DNA to become associated with the yeast cell wall. In the transformation for 18C, Imp4 and Rrp9 I had already been following general advice for high efficiency, such as adding the PEG first as it protects against the high concentration of lithium acetate, and being gentle with the yeast cells after heatshock.

Table 2 shows that all conditions, even the original unoptimized, seemed to provide a transformation efficiency of >6250 transformants per µg of DNA. The similarity between all conditions also indicated that the transformation was already at peak optimality for this strain. Considering these results, it seemed the increased scale was the limiting factor not the procedure. I postulated that perhaps the impact of the heat shock was compromised by the increased

volume of the transformation mix. I tested this by performing a competition scale transformation, splitting the 10.8 ml transformation mix into four 2.7 ml aliquots during the heat shock. This resulted in 2,660,000 transformants, higher than my original target of 500,000, showing that indeed this was the limiting factor. All subsequent experiments using the optimised variation of splitting the transformation mix have a higher number of transformants due to this change. This results in lower noise between wild type variants, and presumably for all variants (**Table 1**)(**Figure 21**). **Figure 20A** shows that the correlation between transformants and variation is not linear, with there being no decrease in variance after ~500,000 transformants. It may be that the protein depletion inherently increases variation, increasing the necessity for a high transformation efficiency. This could explain 18C being less affected by the low transformation efficiency. There may also be transformation efficiency differences between strains however this does not seem to be the limiting factor.

Table 2, Transformation optimisation

Transformation conditions	Scale	Transformation efficiency (Transformants per µg of DNA)	Transformants
Experiments from (Puchta, et al., 2016)			
-	X30	6250	500,000
Unoptimised			
-	X30	~2000	~160,000
Transformation conditions for optimisation			
-	X1	31,875	85,000
2-fold less DNA	X1	24,375	65,000
2-fold more DNA	X1	29,063	77,500
30 minutes at 30C before heat shock	X1	30,187	80,500
Resuspend in LiAc (100mM)	X1	29,819	79,500
Resuspend in SS + DNA	X1	29,063	77,500
Divide transformation mix for heat shock	X30	33,250	2,660,000

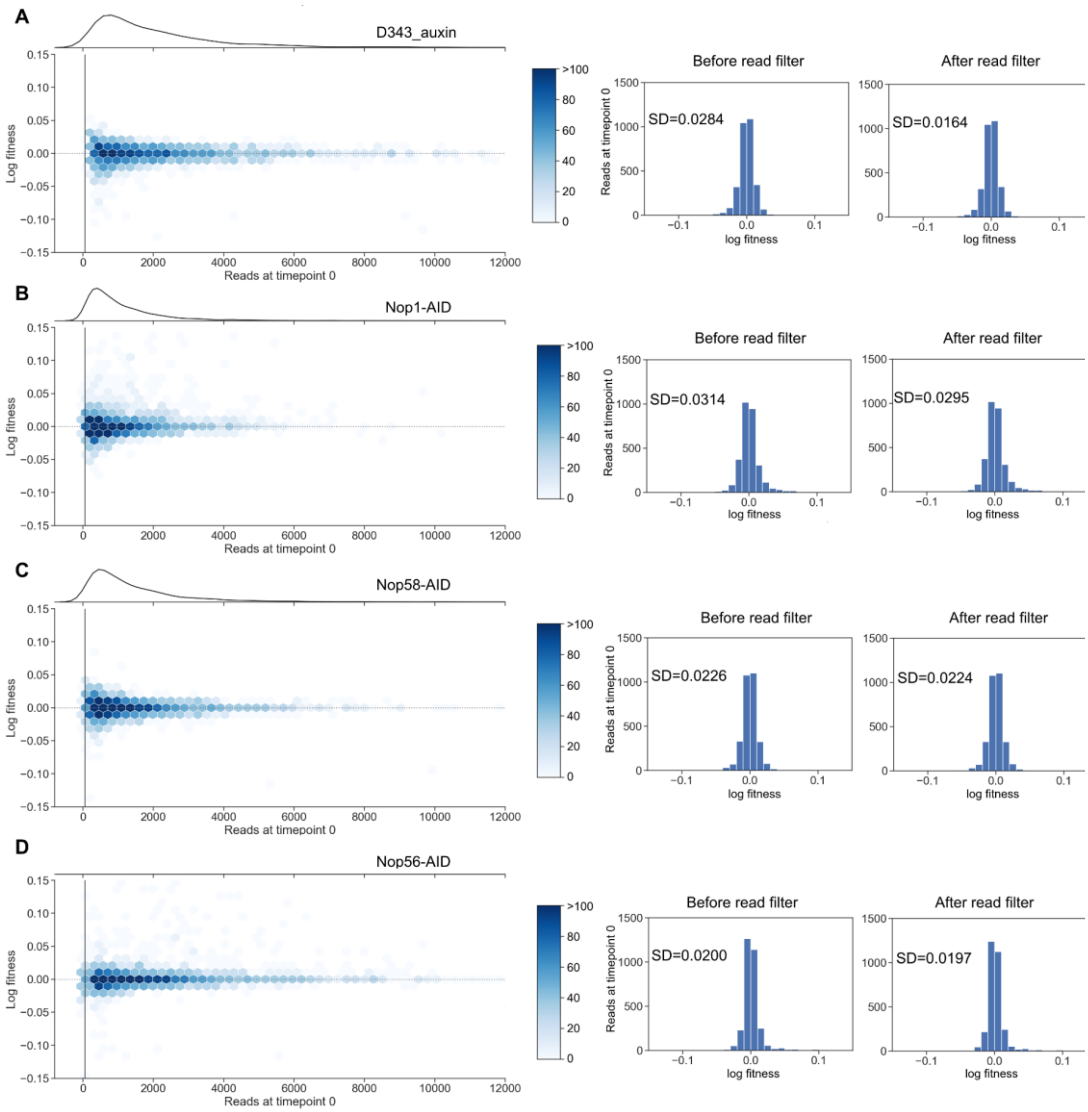


Figure 21, Noise as a function of coverage using wildtype variants, after optimisation

Hexbin density plots for D343_auxin (**A**), Nop1-AID (**B**), Nop58-AID (**C**), Nop56-AID (**D**), visualising noise as function of coverage , with log fitness on the y axis and reads found at the first time point for each barcode on the x axis. A line indicates the read count filter, and a darker shade of blue indicates more barcodes per hexagon. The same data is also plotted in a distribution plot for each experiment before and after read count filtering with the standard deviation value. The values shown have been normalised to 0 by subtracting the median log fitness value obtained after read count filtering.

3.5.2 Accurately calculating log fitness with different growth rates

The calculation of log fitness values from the barcode data is detailed in the Materials and Methods section of this thesis. In (Puchta, et al., 2016) the t variable was days post addition of glucose, as the only comparisons necessary were between replicates, which grew as the same rate. When utilising the assay to study genetic interactions however, I found the conditions resulted in variable growth rates. As expected culturing in non-physiological temperatures and downregulating essential proteins caused growth defects. When calculating fitness for these experiments with the intention of comparison, using days as the t variable was no longer suitable. Using days would lead to underestimation of fitness effects in less permissive conditions. Instead I used OD_{600} values obtained when splitting cultures and harvesting time points to calculate the number of generations that had occurred. Using the number of divisions rather days should allow a more fair comparison between experiments carried out at different growth rates. Assuming exponential growth (splitting the culture attempted to keep OD_{600} <1) the number of generations that had occurred between two OD_{600} readings was calculated by 'n = log2(Nt/N0)'. Where n is the number of divisions, N_0 is the OD_{600} reading at the initial measurement and N_t is the OD_{600} reading at the subsequent measurement. These values were used to approximate how many generations had occurred at each time point. The number of generations was then substituted for daysin the calculation of log fitness. This facilitated the comparison of experiments, as the consideration of divisions when calculating log fitness estimates should compensate for the difference in growth rates.

4 INVESTIGATING GENE-GENE INTERACTIONS WITHIN THE FITNESS LANDSCAPE OF U3 SNORNA

4.1 Characterising gene-gene interactions

The way mutations influence phenotype is a key question. Mutations do not however exist independent of one another, and as such genetic interactions can exist between them. These can occur in *cis* (within one gene, intragenic epistasis) or *trans* (between two genes, intergenic epistasis, gene-gene interaction). This makes predictions more complex, as the genetic background in which mutations occur must be considered. It is termed a genetic interaction, or epistasis, when the observed effect of mutations is different than would be expected under an additive model. Combinations of neutral, deleterious and beneficial mutations can range from negative to positive epistasis, potentially even creating novel functions.

Genetic interactions have been observed at a genome-scale using the wealth of molecular tools available when working with yeast. Systematically analysing data from gene deletion and hypomorphic alleles has allowed the 'genetic landscape' of a cell to be derived (Tong, et al., 2004 , Costanzo, et al., 2010). Genetic interactions were abundant between 5.4 million gene-gene pairs, clustering into established pathways. Many genetic interactions were able to expose the redundancy that gives the eukaryotic genome such robustness against genetic perturbation. In another study it was found that interactions between 87 mutation pairs in more than 500 closely related tRNAs varied from negative to positive depending on the genetic background (Domingo, et al., 2018). The prevalence of interactions highlights the importance of understanding the molecular mechanism behind them. Different mechanisms, found both found empirically and proposed theoretically, explain how these genetic interactions manifest in a molecular context (Lehner, 2011). An example of negative intermolecular epistasis is in duplicated genes where the loss of one is tolerated, but both leads to loss of function. While two individually detrimental mutations within a protein coding gene can combine to allow

interactions with novel substrate, an example of positive intramolecular epistasis. Finding what particular scenarios exist and what factors influence them is vital if ever to predict them. The use of *in silico* methods such as machine learning are vital in mammalian cells, as the same systematic genetic interaction screens used in yeast don't exist (Madhukar, et al., 2015， Boucher & Jenna, 2013). Identifying how genetic interactions lead to disease phenotypes is key to effective therapeutic strategies. The degree to which genes were sensitive to gene-gene interactions correlated with their sensitivity to chemical-gene interactions, having implications for drug targets (Costanzo, et al., 2010).

4.1.1 Characterising intragenic epistasis

The use of deep mutational scanning has revealed the patterns of intragenic epistasis within genes. The systematic and massively parallel nature of these studies provides a complete picture of interactions within a gene by providing a quantitative comparison between thousands of genotypes. This has exposed the presence of both positive and negative epistasis, the degree to which is being heavily dependent on the gene, with some presenting very few cases of positive epistasis (Bendixsen, et al., 2017). The gene level resolution of these studies enables more biochemical interpretation than at a genome scale. This enables molecular mechanisms to be prescribed to the genetic interactions observed. The patterns of epistasis can even be used to predict protein structures (Schmiedel & Lehner, 2018， Rollins, et al., 2018), or improve RNA structures (Puchta, et al., 2016) by providing constraints based on compensatory positive epistasis between base pairs.

Using fitness (growth relative to a reference) as the phenotypic readout in deep mutational scanning assays has enabled the fitness landscape and the

network of intragenic epistatic interactions to be derived for a number of genes. Examples include the proteins Hsp90 (Bank, et al., 2015) (Hietpas, et al., 2011), APH(3')II (Melnikov, et al., 2014), BRCA1/DBR1 (Findlay, et al., 2014), GFP (Sarkisyan, et al., 2016), and RNAs U3 snoRNA (Puchta, et al., 2016) and tRNA-Arg(CUU) (Li, et al., 2016). The seeming propensity towards negative epistasis within these genes suggests a concave fitness landscape. In this scenario a wild type sequence is well adapted, located in sequence space atop a fitness plateau. Any deviation from this sequence ventures into the surrounding low fitness valley, whose steepness increases upon further dislocation. This increasing steepness is reflected in the successive increase in strength of negative fitness effects as a factor of mutation number. There was also considerable mutational robustness observed, with many mutations having a neutral impact upon fitness. Many of these observations aligned with functional expectations, with high levels of epistasis involving functionally important regions and neutral genetic backgrounds being enriched in robust regions.

4.1.2 Characterising intergenic epistasis

A less explored field compared to intragenic epistasis is the systematic mapping of intergenic epistasis. While genomic level studies have identified its abundance as a large network, the mechanisms and finer details of these gene-gene interactions have not been characterised. The use of substitutions in systematic mutational approaches would also provide results more representative of physiological occurrences than the large scale deletion studies.

One finding in the genetic landscape of a cell was that physically interacting molecules were prone to genetic interactions (Costanzo, et al., 2010). An

example of this was finely mapped using mutant libraries of two proteins that physically interacted (Diss & Lehner, 2018). Genetic interactions were found to match functional expectations, with disruptions to protein complex formation being most common. These were found to either alter thermodynamics beyond a threshold or directly affect structural interactions between proximal residues. This study highlighted how predictable molecular mechanisms can explain the observed impact of the genetic background.

4.1.3 Investigating intergenic epistasis between U3 snoRNA and its interacting proteins

In this chapter I detail the identification of genetic interactions between the U3 snoRNA and a panel of interacting protein genes. I achieved this by combining gene replacement via a U3 snoRNA mutant library and hypomorphic levels of interacting proteins. The downregulation of protein levels represents mutations that reduce expression such as those in regulatory elements. It should also be a good approximation of hypomorphic mutations that impair protein function. As both U3 snoRNA and the interacting proteins are essential using null alleles would not be possible. Instead the use of mutations in U3 snoRNA ranging from neutral to lethal combined with hypomorphic protein levels via the AID system gives scope for interactions to be detected. For each experiment where a protein is downregulated the fitness landscape of U3 snoRNA under these conditions was obtained. A genetic interaction was scored when a genotype under hypomorphic conditions has a different fitness impact compared to the wild-type strain.

The use mutations instead of knockouts and fitness as the selection criteria gives an important *in vivo* context to the results. Due to the multiple protein and RNA interactions essential to U3 snoRNA's function it provides the

opportunity to look at several molecular mechanisms. This can vary from directly disrupting these interactions, to perturbing the secondary structure of U3 snoRNA dictated by structural RNA elements and motifs. In **Figure 22** I have outlined a case of synthetic lethality whereby both a mutation and hypomorphic protein levels are tolerated separately, but together are deleterious. The breadth of genotypes the library also enables to study of specific mutations, such as those that only disrupt G-C pairs, or those that close internal loops. The abundance of interactions in *cis* has already been identified in U3 snoRNA (Puchta, et al., 2016) and I expected interactions to also be abundant in *trans*, as found between two proteins in (Diss & Lehner, 2018).

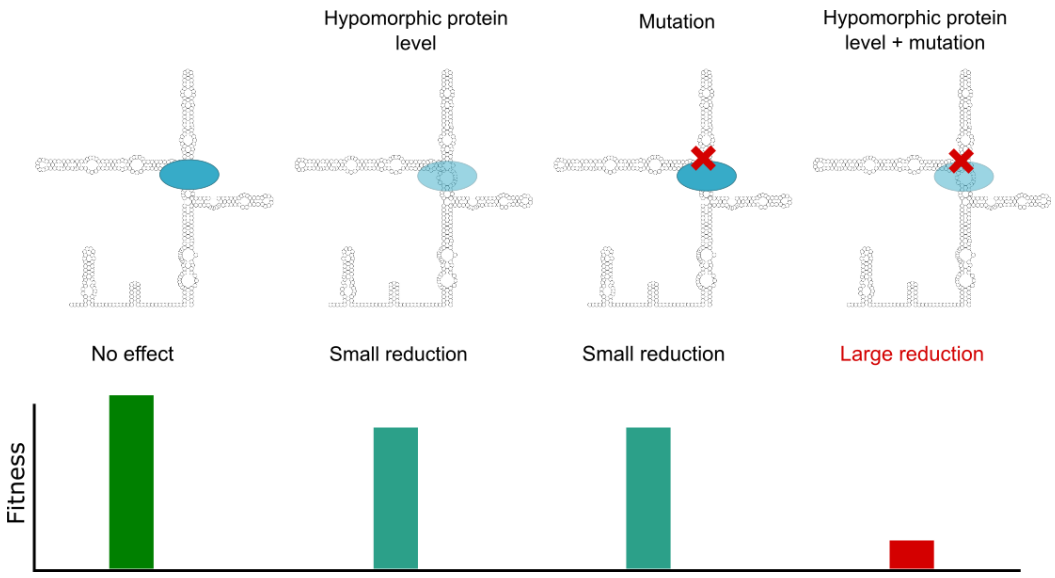


Figure 22, Synthetic leathality

Illustrating a potential example of synthetic lethality. Hypomorphic protein expression levels or a mutation has a small reduction on fitness separately, but combined has a large effect.

4.1.4 Proteins of interest

The proteins chosen for downregulation represent key partners of U3, essential for its role in pre-rRNA processing. These are Nop1, Imp3, Imp4, Rrp9, Nop56, Nop58, Mpp10, Snu13, Dhr1 and the non U3 interacting control Emw1.

In **Figure 23** I have produced a cartoon of U3 snoRNA and the aforementioned proteins using cryo-EM data of the SSU processome from (Barandun, et al., 2017). At this stage of pre-rRNA processing U3 snoRNA has bound the pre-rRNA but is prior to cleavage. The U3 RNP proteins (Nop1, Nop56, Nop58, Snu13, Rrp9) can be seen tightly clustered around the stem of U3 snoRNA. This stem contains the two pairs of juxtaposed box motifs that constitute the protein binding k-turn motifs. Nop56 and Nop58 form a heterodimer connecting the two pairs of box motifs. The proteins that constitute the U3 RNP apart from Rrp9 are conserved throughout all box C/D snoRNAs. Snu13 binds to the k-turn motif of U3 snoRNA near the box B/C motif which undergoes conformation changes to then facilitate the subsequent binding of Rrp9. Nop1 acts a methyl-transferase in other snoRNAs with this role being usurped for a structural role perhaps as U3 snoRNA transitioned away from a more traditional snoRNA function. Within the SSU processome the two copies of Nop1 in the U3 snoRNP are binding platforms for five proteins (Sas10, Utp24, Utp11, Fcf2, and Bud21). The heterodimer of Nop56 and Nop58 are vital for proper organisation within the SSU processome, while Nop58 is also important for U3 snoRNA stability. Imp3, Imp4 and Mpp10 make up the Mpp10 complex, and are involved in both duplex formation and disassembly between U3 snoRNA and the pre-rRNA. Dhr1 is a DEAH-box RNA helicase, partially responsible for the disassociation of U3 snoRNA from the pre-rRNA. All selected proteins directly interact with U3 snoRNA in some capacity. This creates the expectation that during downregulation intergenic interactions should be enriched in areas of the U3 snoRNA proximal to the protein.

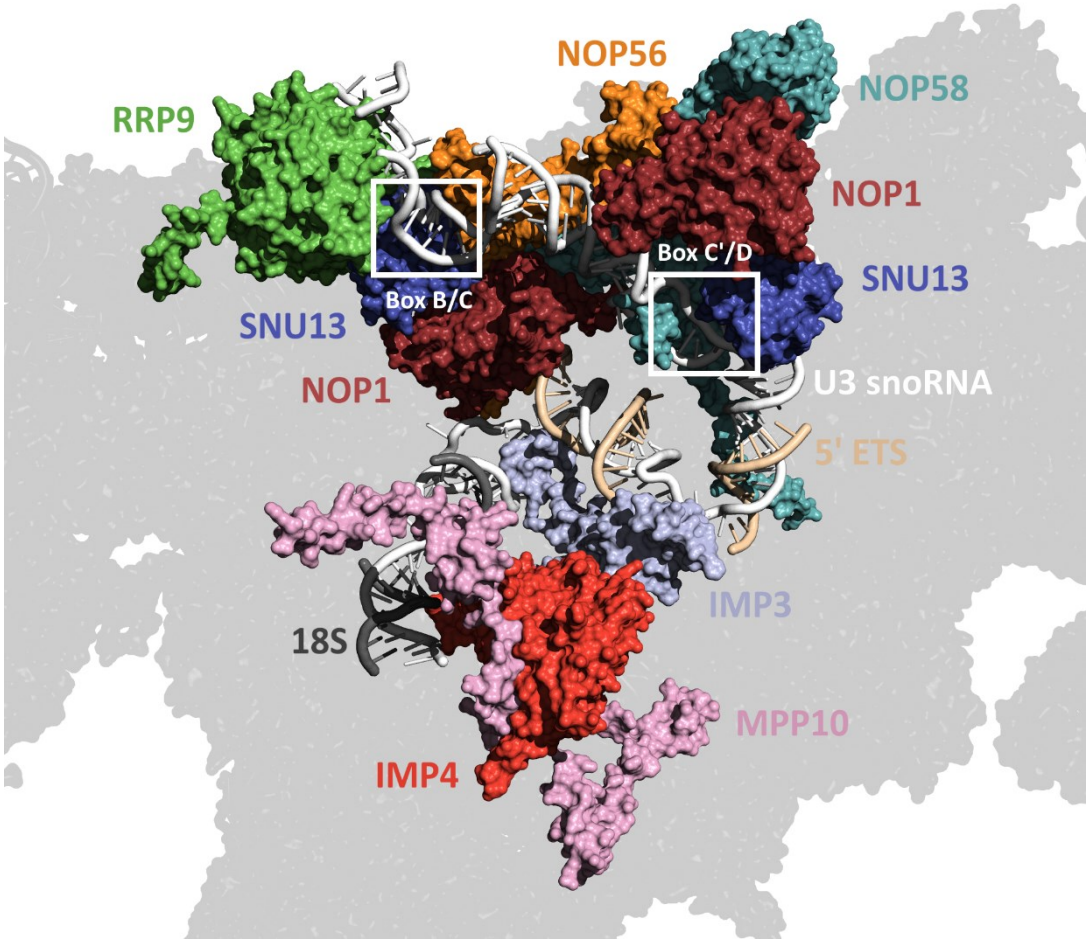


Figure 23, Cryo-EM map of the U3 RNP and Mpp10 complex within the SSU processome

A cryo-EM derived cartoon representation of the SSU processome, produced using data from (Barandun, et al., 2017). The SSU processome is greyed out with the following highlighted: U3 snoRNA (white), the regions of 18S (black) and 5' ETS (wheat) that base pair with U3 snoRNA, Nop1 (rouge), Imp3 (lilac), Imp4 (red), Rrp9 (green), Mpp10 (pink), Snu13 (blue), Nop56 (orange) and Nop58 (teal). The B/C and C'/D box motifs are indicated by white boxes.

There may be interactions common to all proteins as well as those that are protein specific. Additionally, as these proteins also have essential roles within other box C/D snoRNPs that direct pre-rRNA modifications, their downregulation may have downstream implications on the fitness landscape of U3 snoRNA. The pleiotropy of genes such as this contributes to the complexity of characterising genetic interactions.

4.1.5 Competition experiment using Auxin inducible degron strains

As detailed in the previous chapter, I used the auxin inducible degron system to downregulate the proteins of interest. This reduces protein levels as a function of auxin concentration. I completed the completion experiment in 350μm of NAA (auxin), where western blotting in the Dhr1-AID strain indicated protein levels were reduced by ~80%. The competition experiment was carried out (detailed in Materials and Methods section of this thesis) with the addition of auxin coinciding with the addition of glucose, just after collecting the first time point (**Figure 24A**). This meant both U3 snoRNA and the protein of interest should begin to be downregulated at the same time. This was to reduce any potential dominant negative effects that could manifest if the protein was downregulated prior to genomic U3 snoRNA being repressed. This could alter the distribution of the starting pool in a biased manner. In the future completing the competition in galactose conditions to intentionally look for dominant negative genotypes would be an interesting follow-up experiment.

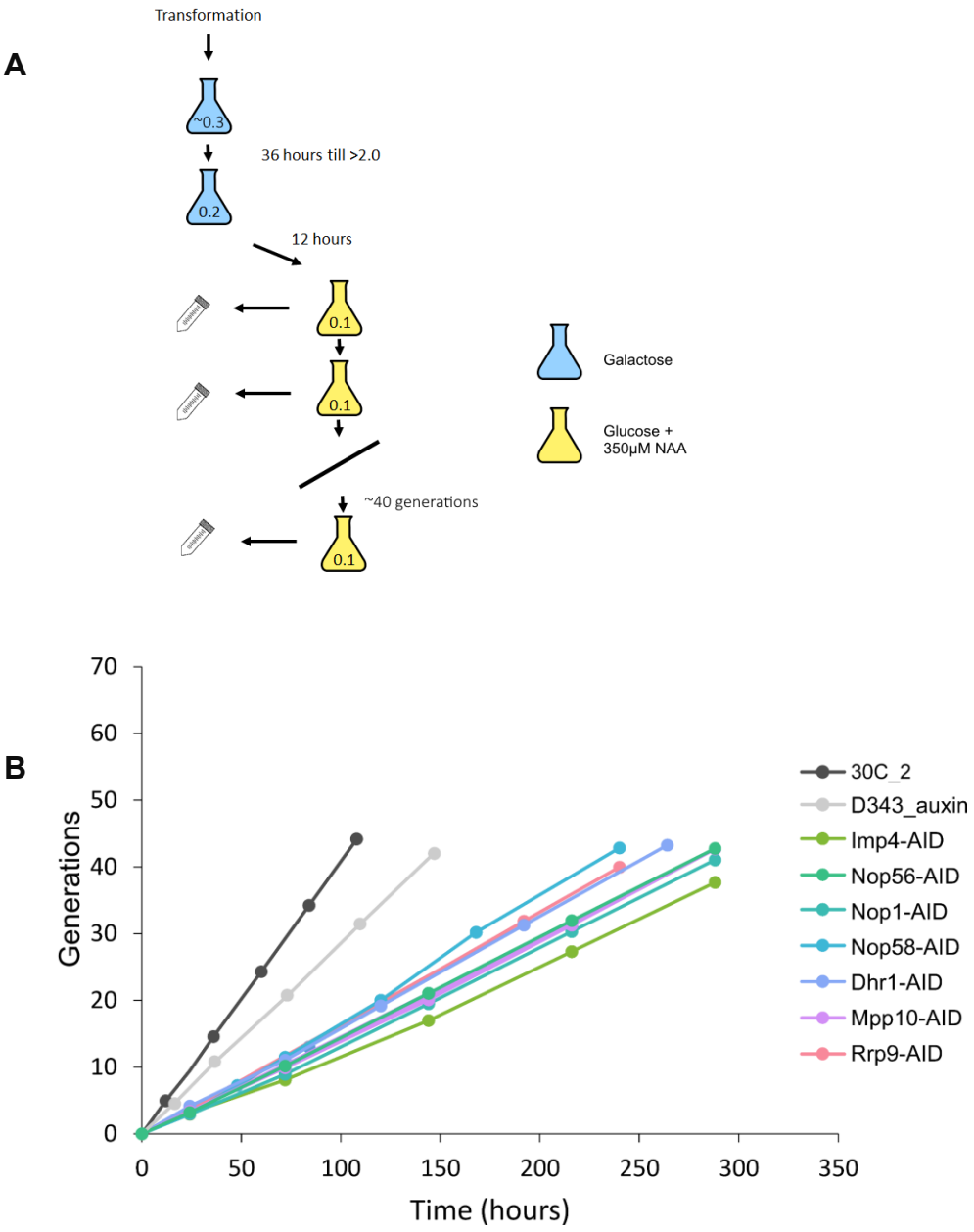


Figure 24, Competition experiment with AID strains

A. Schematic detailing the steps of the competition experiment. After transformation the population is cultured for 36 hours, with the population reaching an OD_{600} value of >2.0 if a high transformation efficiency is achieved (quantified by plating). This is then split to 0.2 and cultured for 12 hours to obtain a population of mainly transformed cells. This is then split to 0.1 and cultured in glucose media and maintained in log growth phase by periodically splitting to 0.1 OD_{600} and harvesting time points. The number of times the culture was split is truncated with a diagonal line. The number within the flasks indicates the OD_{600} measurements the culture was split to.
B. Line chart illustrating the number of generations occurring over time (hours) for the different strains over the course of the respective competition experiment. All were performed in auxin apart from 30C_2.

Due to the slower growth in non-permissive conditions (**Figure 24B**) the AID strains were split less often than the wild type strain. While splitting to maintain log growth OD_{600} measurements were taken every time culture was split to calculate the number of generations that had occurred since the last measurement. Yeast were harvested at each of these events. The experiment was concluded after exceeding ~40 generations, with the six time points chosen to be sequenced being proximal to 0, 5,10,20,30 and 40 generation. The exact number of estimated generations is shown in **Table 3**. These generation milestones were chosen to make them comparable to the time points from the wild type strain experiment and to each other. I decided to cluster time points slightly towards the beginning of the experiment as this had been predicted to improve statistical power (Matuszewski, et al., 2016). This would also ensure any deviations between highly deleterious genotypes could be more highly resolved.

Mpp10-AID appears to grow at a similar rate to the other strains, which is contrary to its slower growth while being cultured (**Figure 16**). To look for potential contamination of another strain, I genotyped all time points of the Mpp10 experiment which showed no contamination (not shown). This could be the result of strong gene-gene interactions causing some genotypes to outcompete the wild-type genotypes.

The yeast extracts from selected time points then underwent plasmid extraction. The plasmids were used as a template to add illumina sequencing adapters either side of the molecular barcode sequence. These adapters were designed to ensure the first base sequenced by the illumine platform was the barcode sequence to maximise library diversity.

Table 3, AID strain time points

Experiment Name	Time point	Time (hours)	Divisions
30C_1	1	0	0
	2	36	15
	3	60	25.1
	4	84	35
	5	108	44.9
30C_2	1	0	0
	2	12	4.9
	3	24	9.4
	4	36	14.5
	5	60	24.3
	6	84	34.2
	7	108	44.14
Imp4-AID	1	0	0
	2	24	3.1
	3	72	8
	4	144	16.9
	5	216	27.2
	6	288	37.6
Rrp9-AID	1	0	0
	2	24	3.2
	3	48	7.2
	4	120	19.5
	5	192	31.8
	6	240	39.9
D343_auxin	1	0	0
	2	16	4.5
	3	36	10.7
	4	72	20.7
	5	108	31.5
	6	144	42
Nop1-AID	1	0	0
	2	24	2.9
	3	72	8.9
	4	144	19.4
	5	216	30.3
	6	288	41
Nop58-AID	1	0	0
	2	24	3
	3	72	11.4
	4	120	20
	5	168	29
	6	240	42.8
Mpp10-AID	1	0	0
	2	24	3.4
	3	72	9.8
	4	144	20
	5	216	31.2
	6	288	42.6
Nop56-AID	1	0	0
	2	24	3.2
	3	72	10.1
	4	144	21
	5	216	3139
	6	288	42.7

4.2 Fitness landscape of U3 snoRNA with hypomorphic protein alleles

The sequencing results are sorted into their respective time points using index sequences assigned during library creation. The number of times a barcode occurs in each time point is counted, giving an estimation of its copy number in the yeast population. When extracting the barcodes initially, the reads were processed using bedtools, bash and awk to find the sequence flanking the barcode. This 3' flanking sequence should be identical in all reads, with the remaining sequence presumed to be barcode sequence. These barcode sequences were then processed using an awk script that attempted to uniquely match them to a list of filtered barcodes obtained in the MiSeq experiment, allowing for at most two mismatches. This would then give a list of high confidence barcodes. I streamlined this somewhat by using blast, directly searching the reads using a database of the filtered barcodes. The stringency used again allowed for at most two mismatches. This was advantageous as it was only one step, omitting the exact matching of the 3' flanking sequence, which would ignore reads that had sequencing errors in this region. Exponential decay curves were then fitted to each barcode's count data, used to measure the logarithm of relative fitness for that particular genotype. These fitness values are then normalised using the wild type variants present, adjusting neutral mutations to a log fitness value of 0. The final processed data for each experiment is a list of ~24,000 U3 snoRNA genotypes with assigned log fitness values.

I was able to complete the competition experiment in 350μm of auxin for six strains (Imp4-AID, Rrp9-AID, Nop1-AID, Nop58-AID, Mpp10-AID, Nop56-AID), as well as the D343 control strain. Although I did not complete the experiment with Emw1, the non U3 interacting control, the D343 control strain in auxin

should serve as an adequate control. At the time of this thesis however, I was not yet able to analyse the data from the Mpp10-AID and Nop56-AID experiments. I will also be using two control datasets from (Puchta, et al., 2016), “Small_1_30C_Glu” and “Small_3_30C_Glu” which I refer to as “30C_1” and “30C_2” respectively. The log fitness values from the AID experiments ultimately represent the impact of different U3 snoRNA genotypes in the background of hypomorphic alleles.

4.2.1 Distribution of log fitness effects

In **Figure 25** I have plotted all genotypes and just single mutant genotypes for: the two 30°C repeats, D343_auxin and the AID strain experiments. The log fitness distribution of all genotypes follows a negatively skewed distribution, maintained across all experiments. This pattern of fitness effects has been observed in multiple similar deep mutational scanning experiments. This indicates the majority of mutations are either neutral or deleterious, with beneficial mutations being rare (Bendixsen, et al., 2017 , Sanjuán, et al., 2004). The peak around 0 log fitness reflects the large regions of U3 snoRNA that exhibit mutational robustness. The wider, lower log fitness peak represents mutations that affect U3’s function. These genotypes could be deleterious due to a number of factors such as reducing U3 snoRNA’s stability, disrupting protein binding or effecting pre-rRNA duplex formation. There is however a broader range of fitness effects with U3 snoRNA than with other some fitness experiments, as many genotypes can impair U3’s function without completing ablating it.

There is variation in the distribution of log fitness effects between the two 30°C repeats and D343_auxin that is greatly reduced when calculating f_i and p_i. This

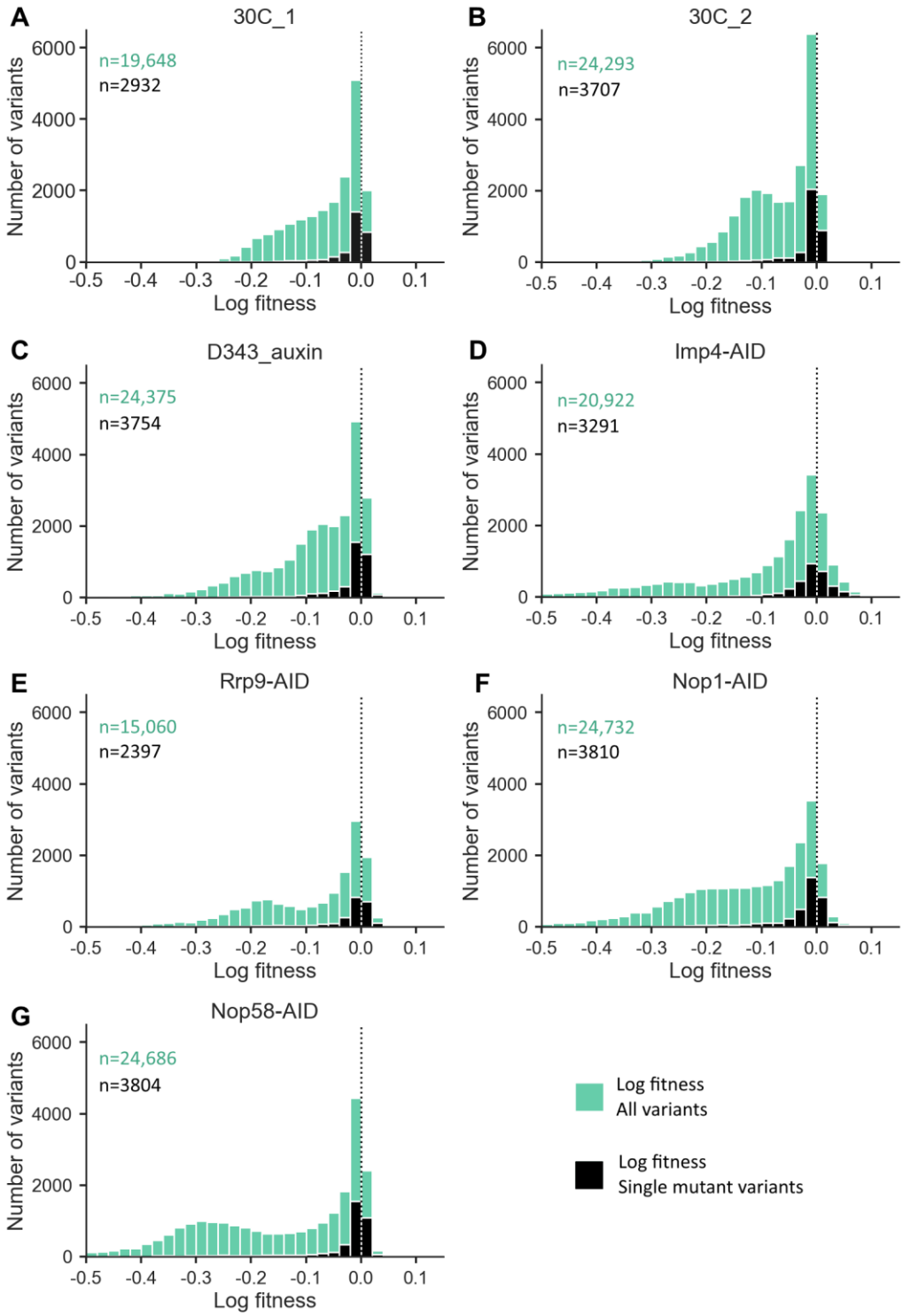


Figure 25, Distribution of log fitness effects in the AID strains

The number of genotypes detected in an experiment (teal), as well as just single mutant genotypes (black), plotted in 0.02 log fitness bins. The number of variants that constitute each group is indicated in its respective colour.

is partially due to the difference in variants that are detected between experiments, with the number of variants detected listed in **Figure 25**. The degree of fitness impacts however are distributed over a very similar range, with the tail of the peak containing deleterious genotypes ending around -0.3 (**Figure 25 A-C**). There is also a similar proportion of genotypes in the lower fitness peak, with 63% of 30C_1, 65% of 30C_2 and 67% of D343_auxin genotypes being deleterious (<-0.02 log fitness). The single mutant genotype distributions are more similar between the three experiments. The bimodal distribution is not visible on the scale of **Figure 25** as only a small number of genotypes are deleterious.

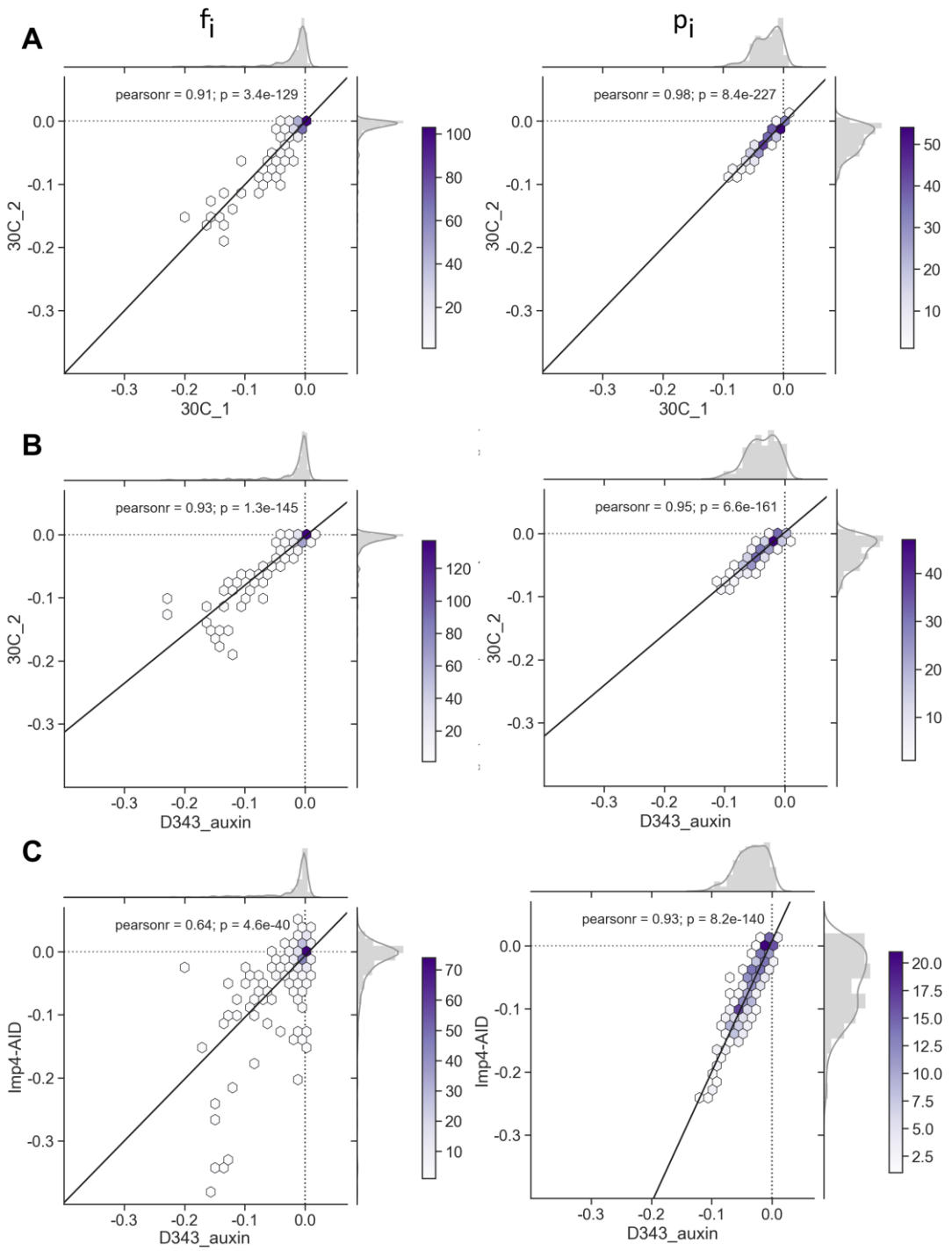
The log fitness effects in the AID strain experiments are all distributed over a larger area, with the tail of the peak for Imp4-AID, Nop1-AID and Nop58-AID ending around -0.5. Rrp9-AID appears the least effected by this shift (**Figure 25E**), with a similar proportion of deleterious genotypes as the controls (65%). Imp4-AID also has a similar proportion (65%), while Nop1-AID and Nop58-AID have a higher proportion at 76% and 72% respectively. This pattern suggests a higher prevalence of genotypes that were previously neutral exhibiting a negative fitness effect in the Nop1-AID and Nop58-AID experiments. This could result from more genotypes breaching thresholds that are lowered in the background of a hypomorphic mutation for the respective protein. With Imp4-AID and Rrp9-AID the pattern seems to indicate for the most part previously negative genotypes have simply become more so. The variation between the distributions show that within the overall shift perhaps caused by downregulating an essential protein, there are gene specific interactions. The detrimental peak in the Nop58-AID experiment for example shows a large proportion of genotypes around -0.3 (**Figure 25G**), indicating many genotypes and likely many positions are very sensitive to mutation.

The inherent technical noise for each experiment can also be observed, similarly to when looking at the distribution of wild type genotypes (**Figure 21**). For Imp4-AID (**Figure 25D**) for example there appears to be genotypes with positive fitness effects. This can be explained however by the standard deviation observed for the wild type genotypes, which encompasses the 'positive' genotypes. The distribution of wild type variants can be seen in **Figure 19D**, covering the same area as the seemingly positive genotypes seen in **Figure 25D**.

4.2.2 Correlation between experiments

Comparing the f_i and p_i values from various experiments reveals any positions that deviate from the control. In **Figure 26** I have plotted fi values and p_i values of different experiments against one another to expose genetic interactions, comparing two 30°C repeats from (Puchta, et al., 2016) 30C_1 and 30C_2, comparing 30C_2 to D343_auxin, and comparing D343_auxin to the AID strain experiments. The f_i and p_i values are calculated only with variants that made it past filtering in both experiments. This allows for a fair and accurate comparison of each position of U3 snoRNA, avoiding differences potentially caused by genotypes with strong fitness effects being present in one experiment but not the other. This also reduces the variation seen when looking at the distribution of log fitness effects, as calculating f_i and p_i values takes advantage of the ability to produce averages from the ~3,600 (f_i) / ~24,000 (p_i) log fitness values, significantly reducing technical noise.

The two 30C repeats correlate very well as expected (**Figure 26A)**. The slight variation between low fitness f_i values stems from the smaller number of values that constitute f_i (only ~15% of the library are single mutant variants) and an


A
f_i
p_i
pearsonr = 0.91; p = 3.4e-129
pearsonr = 0.98; p = 8.4e-227
30C_2
30C_1
B
pearsonr = 0.93; p = 1.3e-145
pearsonr = 0.95; p = 6.6e-161
D343_auxin
C
pearsonr = 0.64; p = 4.6e-40
pearsonr = 0.93; p = 8.2e-140
Imp4-AID

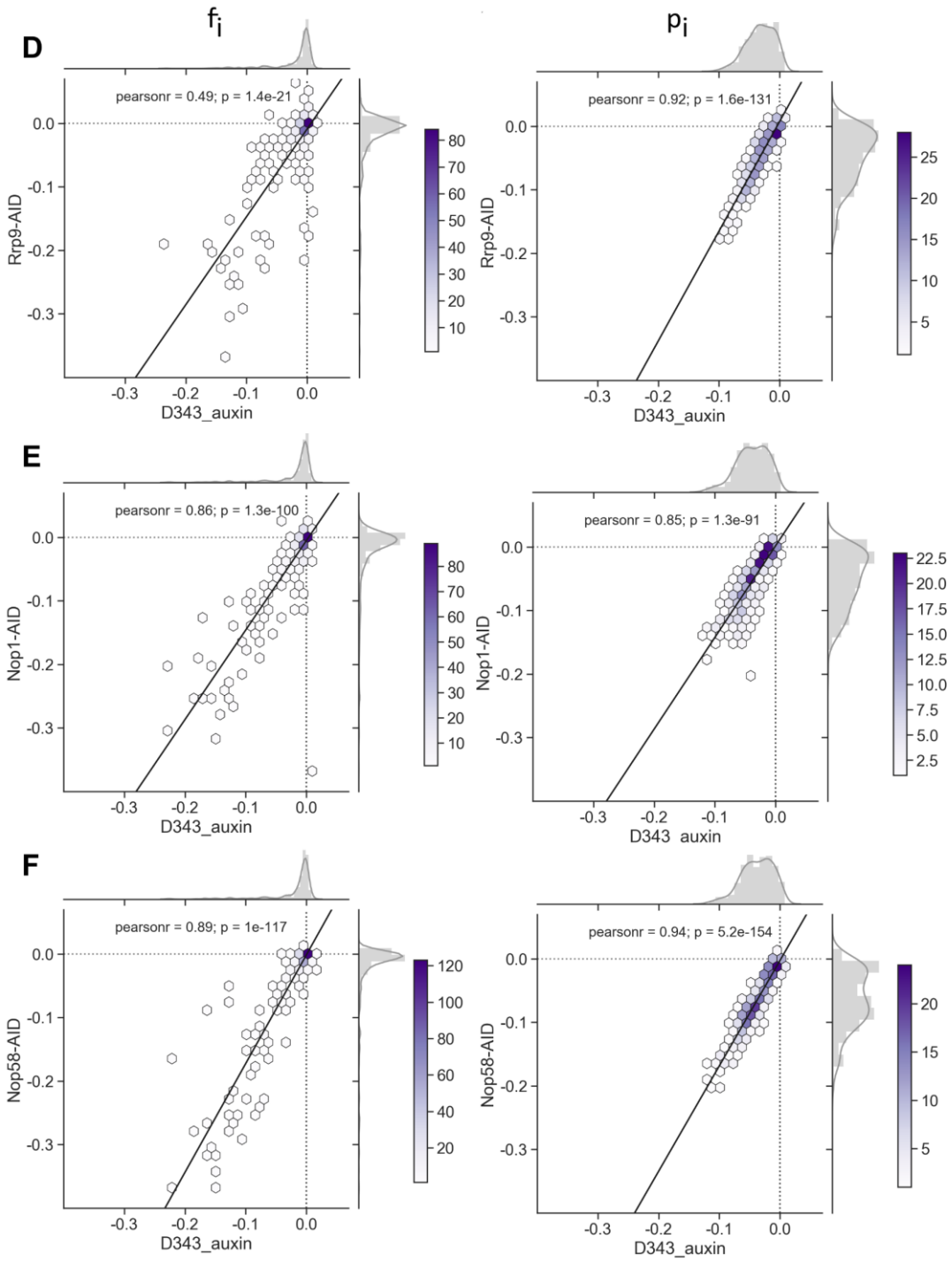


Figure 26, Correlating the f_i and p_i values from AID strains

Hexbin density plots between f_i (left column) and p_i (right column) values for different experiments (**A-F**), with regression line and statistics, and distribution plots on the axes. A scale bar for each plot indicates the number of values that constitute each hexagon, with white indicating one value and a darker shade of purple indicating a higher number.

increased variation when assigning low log fitness effects. The high number of values used when calculating p_i compensates for this. There does appear to be a slight auxin dependent impact (**Figure 26B**), but the effect is minimal compared to the large differences seen in the AID strains. This may be a result of auxin, but could also be a result of the higher sequencing depth. The high number of reads at the first time point gives a more accurate fitness estimation, but also gives barcodes with a relatively low copy number a higher potential to present a large decrease in reads and thus be assigned a lower log fitness value.

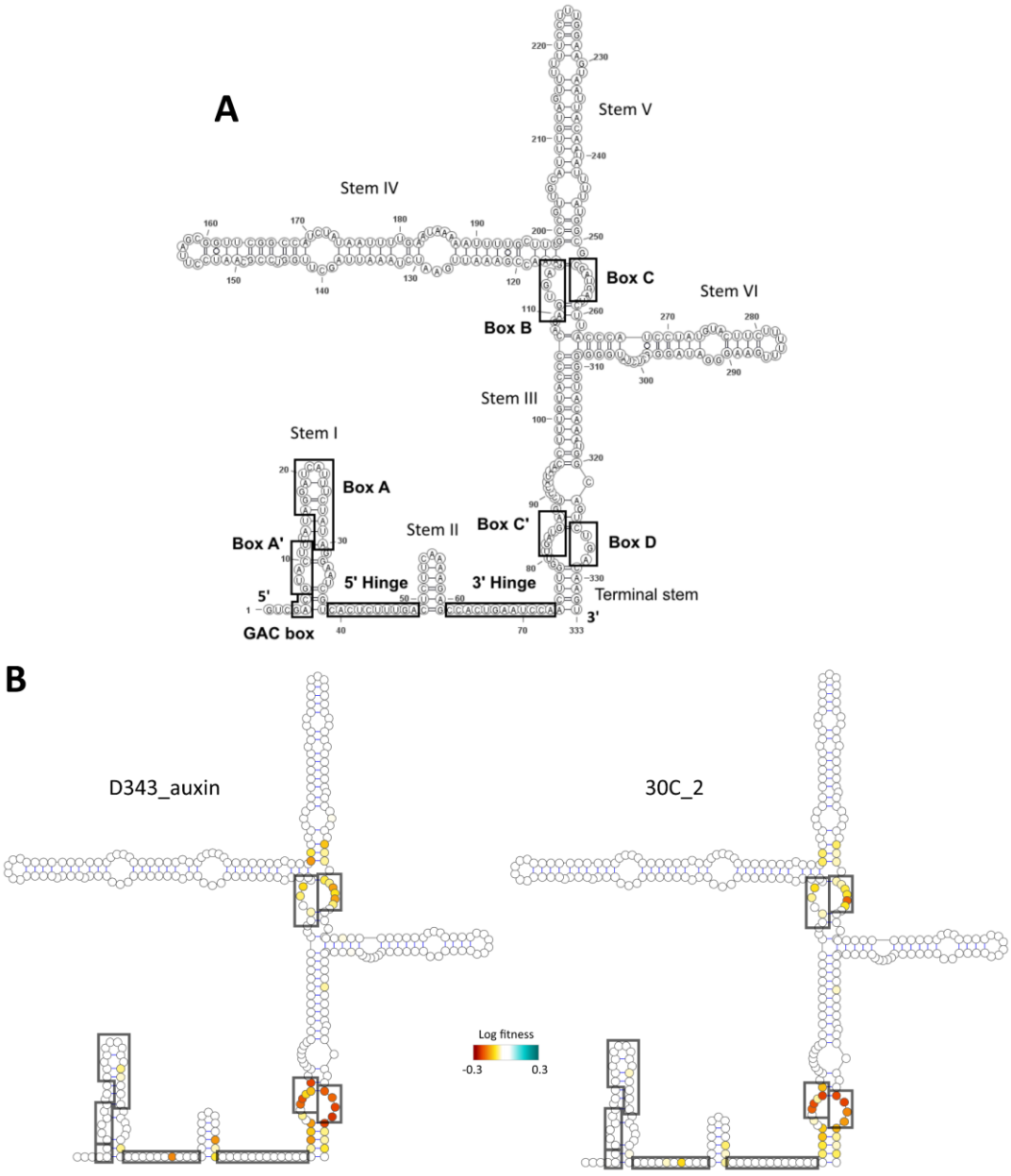
The AID strains deviate considerably in a gene specific manner from the D343_auxin experiment, with many f_i values being drastically different. These represent positions that exhibit intergenic epistasis, deviating from the fitness impact incurred when mutated in a wild type protein expression background (**Figure 26C-F**). Major deviations in f_i values can be found in both positions already sensitive to mutation and neutral positions, showing the extensive robustness to single mutations is reduced at certain positions. The majority of neutral mutations remain unaffected however, with the highest density hex clusters being found at near 0 log fitness. These are positions that reside in regions of U3 snoRNA that still exhibit genetic robustness, able to tolerate mutations even with hypomorphic mutations in the proteins of interest. This same pattern is seen with p_i values, with the majority of near 0 log fitness positions being unaffected.

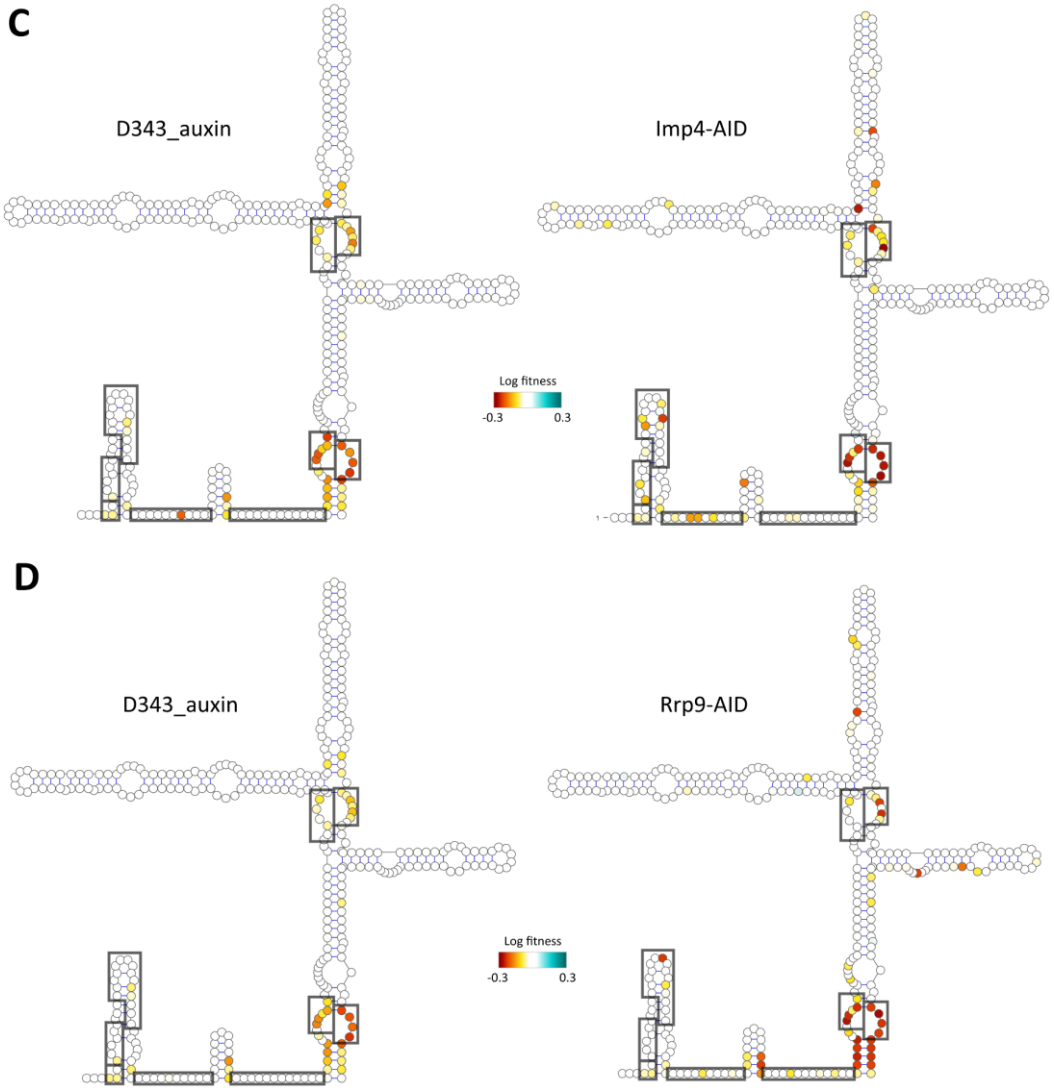
There is an overall negative fitness shift in the AID strain experiments, which nonetheless preserves correlation, highlighting the uniformity of the shift. The preservation of correlation seems to indicate already deleterious genotypes are extenuated by intergenic epistasis, with fewer cases of detrimental fitness effects being found in previously neutral genotypes.

4.3 Mutations in a wild type U3 snoRNA genetic background

From looking the distribution of fitness effects between the different experiments there appears to be strong gene-gene interactions between U3 snoRNA and its interacting proteins. To look from a more functional perspective, I calculated f_i values. These are constituted of only mutations in a wild type background, exposing the positions most sensitive to mutation. The genotypes present in the library result in every U3 snoRNA position being represented by deletions and multiple substitutions.

Figure 27 represents the distribution of f_i values on the secondary structure of U3 snoRNA for 30C_2 (**Figure 27B**), Imp4-AID (**Figure 27C**), Rrp9-AID (**Figure 27D**), Nop1-AID (**Figure 27E**) and Nop58-AID (**Figure 27F**). The colour scale represents the scale of fitness effect. For each experiment the f_i values from D343_auxin are shown, being calculated using only the barcodes also found in the respective experiment, for an accurate comparison. This can cause the presented D343_auxin f_i values to differ slightly between experiments due to differing transformation efficiencies, and thus a differing proportion of the library being detected. This avoids false positives caused by genotypes with strong fitness effects being present in one experiment but not the other. An example of these can be seen with position 191. When substituted to a C this creates a binding site for RNA degradation factor Nab3, thus being highly deleterious. While genotypes with this substitution exist in both the 30C_2 and D343_auxin experiments, by chance none are detected in both, explaining the neutral fi value for this position. In the Nop1-AID and Nop58-AID experiments however genotypes with this substitution are shared with D343_auxin, explaining its low f_i value.


A
Stem V
Stem IV
Box C
Stem VI
Box B
Stem III
Stem I
Box A
Box A'
Stem II
Box C'
Box D
5' Hinge
3' Hinge
Terminal stem
5'
3'
GAC box
B
D343_auxin
30C_2
Log fitness
-0.3
0.3


C
D343_auxin
Imp4-AID
Log fitness
-0.3
0.3
D
D343_auxin
Rrp9-AID
Log fitness
-0.3
0.3

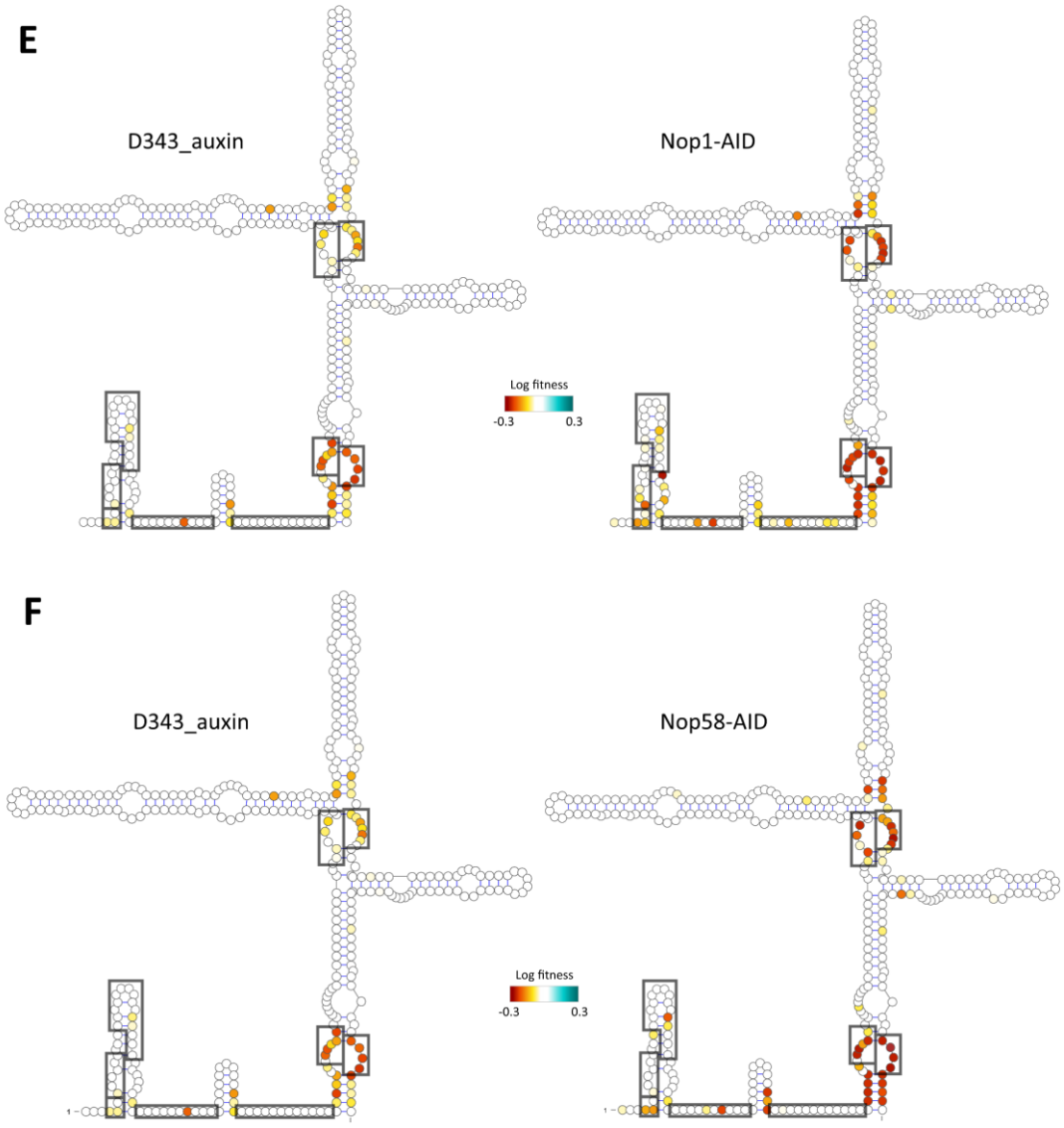


Figure 27, f_i values from AID strains plotted on the secondary structure of U3 snoRNA

A. U3 snoRNA secondary structure with nucleotides labelled, box and hinge regions annotated with black boxes, and nucleotide numbers indicated at 50 nucleotide intervals.

B-F. U3 snoRNA secondary structure with f_i values plotted on a colour scale (indicated by colour bar), from negative fitness effects (red) to positive fitness effects (blue), with white indicating a neutral effects. The f_i values from the D343_auxin experiment are plotted as a control for each, calculated with the same genotypes.

Looking first at the fitness effects in a wildtype protein expression background (**Figure 27B**), the majority of positions of U3 snoRNA display genetic robustness, tolerating single mutations. Positions within the yeast/fungal specific stem IV, V and VI are the most robust. Mutations in the box motifs disrupting the protein binding site k-turn motifs are the most detrimental. This pattern is conserved throughout all experiments. This is followed by mutations in stems, either in the terminal stem at the base of the Box C'/D motif, or at the base of stem II and V. The G-C helices flanking the k-turn motif have been characterised to be important to maintaining a structure capable of protein binding (Klein, et al., 2001). The G-C pairs at the bases of stem II and V are most likely important to maintaining U3 snoRNA's structure. Maintaining stem II between the 3' and 5' hinge is important, as nucleotides 47-63 containing this stem are essential for Mpp10 association (Wormsley, et al., 2001). Maintaining the base of stem V is also important, involved with Snu13 and subsequent protein binding (Granneman, et al., 2002). While changes to the stems are generally of small effect relative to the box motifs in the D343_auxin experiment, the G-C helices in the terminal stem at the base of the box C'/D motif are very deleterious when mutated in combination with reduced levels of Rrp9, Nop1 and Nop58 (**Figure 27D-F**). This could be indicative of the increased importance of the k-turn structural conformation when protein levels are low. With Nop1 the 5' side of the terminal stem is more sensitive to mutation. In the AID strain experiments the degree of fitness effect seen in the box motifs is increased. The mean f_i values for different regions are aggregated in a heatmap (**Figure 28**). The box C, B, C' and particularly box D show large decreases in fitness in the AID strains. The other regions of U3 snoRNA are less affected, maintaining robustness.

Figure 28, Heatmap for f_i values from AID strains in different U3 snoRNA regions

Each row is a U3 snoRNA region (box motifs, hinge regions, GC base pairs and remaining positions), and each column is values from the D343_experiment and protein of interest experiment calculated using the same genotypes. The value is the mean of f_i values in that region, with the colour scale ranging from red for negative log fitness effects, blue for positive log fitness effects and yellow for neutral.

It appears stability may be more of a selective factor with hypomorphic mutations in Nop58. The G-C base pairs at the base of stems II and V have increased sensitivity, and previously tolerant base pairs of stem I and VI now contain deleterious mutations. This conforms with the known role of Nop58 in maintaining C/D box snoRNA stability, with downregulation of Nop58 significantly reducing snoRNA levels (Lafontaine & Tollervey, 1999). G-C base pairs in general are most affected by mutations in the Nop58-AID experiment (**Figure 28**), known to be important to the stability of RNA molecules. The D box, C box, B box, C' box motifs present the strongest genetic interactions with Nop58 compared to other proteins. Not every region is lower however, as the hinge A' box and A box have a high tolerance to mutation relative to other protein's downregulation, perhaps being less important to the stability of U3 snoRNA.

The box B motif initially appears less effected in most experiments (**Figure 28**), however the G and A nucleotides at position 113 and 114 respectively do present fitness effects, especially in the Nop1-AID and Nop58-AID experiments (**Figure 27E-F**). These two nucleotides are important in base pairing with the complementary A and G nucleotides in the box C motif (positions 256 and 257) to form a k-turn motif (Klein, et al., 2001). Therefore, it appears the disruption of the k-turn may be the cause of the observed negative fitness effects. The box B/C motif's function as a protein binding site and subsequent importance for roles such integration into the pre-rRNA processing complex (Granneman S, 2004) may rely on an intact k-turn motif to achieve these roles more than the other positions of the box motifs.

The A box, A' box and hinge region are all complimentary to regions of the pre-rRNA, forming duplexes at the core of the SSU processome. These regions seem particularity sensitive to mutation in the background of Nop1 and Imp4

at hypomorphic levels. These positions are normally robust to single mutation genotypes. This matches expectations for Imp4, with the Mpp10 complex being situated adjacent to the A/A' box stem within the SSU processome (**Figure 23**). These mutations may render this location harder to recognise for the Mpp10 complex, with it being expected that the fitness landscapes for downregulated Mpp10 and Imp3 would exhibit a similar pattern. It might be expected that particular genotypes could confer a positive fitness effect by destabilising the A/A' stem to compensate for the downregulated proteins, but this is not seen while looking at positional effects.

Since I added ~80 single mutations to the library which encompass the first six nucleotides of U3 snoRNA, the impact of mutations at these positions can be determined. Positions 4-6 constitute the GAC box and as expected the most deleterious mutations reside there. The GAC box is necessary for cleavage at site 1 in *Xenopus* (Borovjagin & Gerbi, 2001). Mutations in the GAC box are more deleterious when downregulating Nop1 and Nop58. This is likely due to the overall disruption of U3 snoRNA observed with Nop58-AID, and the increased sensitivity seen in the A/A' stem with Nop1-AID. It might be expected that Imp4-AID shares a similar pattern but this is not the case.

There do appear to be some potentially interesting positions in stems V and VI in the Imp4-AID and Rrp9-AID experiments. On closer examination these are due to the increased variation in these experiments, and in Rrp9-AID's case the low amount of genotypes that passed filtering.

4.3.1 Fitness effect of different substitutions

Calculating f_i values provides the fitness effect of mutating positions with reduced noise, however it hides the effect of individual substitutions. It is not possible to look at every substitution in a wild type genetic background however as the coverage of single mutants in the library does not cover the entirety of them. This can be compensated for by incorporating low-effect double mutants. For each substitution, single mutants are combined with double mutants whose secondary mutation resides in a position with a neutral f_i value. The median log fitness value for each substitution is plotted in **Figure 29**. Inclusion of these low-effect double mutants drastically improves the coverage without altering the results. There are still a small number of missing values marked in grey, especially with the low library size for Rrp9-AID. 30C_2 does not contain any values at positions 1-6 due to being transformed with a library prior to the addition of variants with mutations at these positions.

For the most part all three possible substitutions cause a comparable fitness effect. The pattern is the same as seen when plotting f_i values on the secondary structure of U3 snoRNA (**Figure 27**), with the four box motifs presenting the lowest log fitness values. Imp4-AID and Rrp9-AID appear to have examples of heterogeneity. Upon investigation of these substitutions however the low number of genotypes and inherent variation is the cause, with the positive fitness effects also being explained by this. These two experiments require the reduction in noise afforded by incorporating more of the library in order to be reliable. With human U3 snoRNA the effect of mutating the regions surrounding box B/C motif including stem V and the subsequent effect on hU3-55k (Rrp9) binding has been extensively characterised (Granneman, et al., 2002). If there were more genotypes present in the Rrp9-AID experiment, it would have been interesting to note the effect of equivalent genotypes.

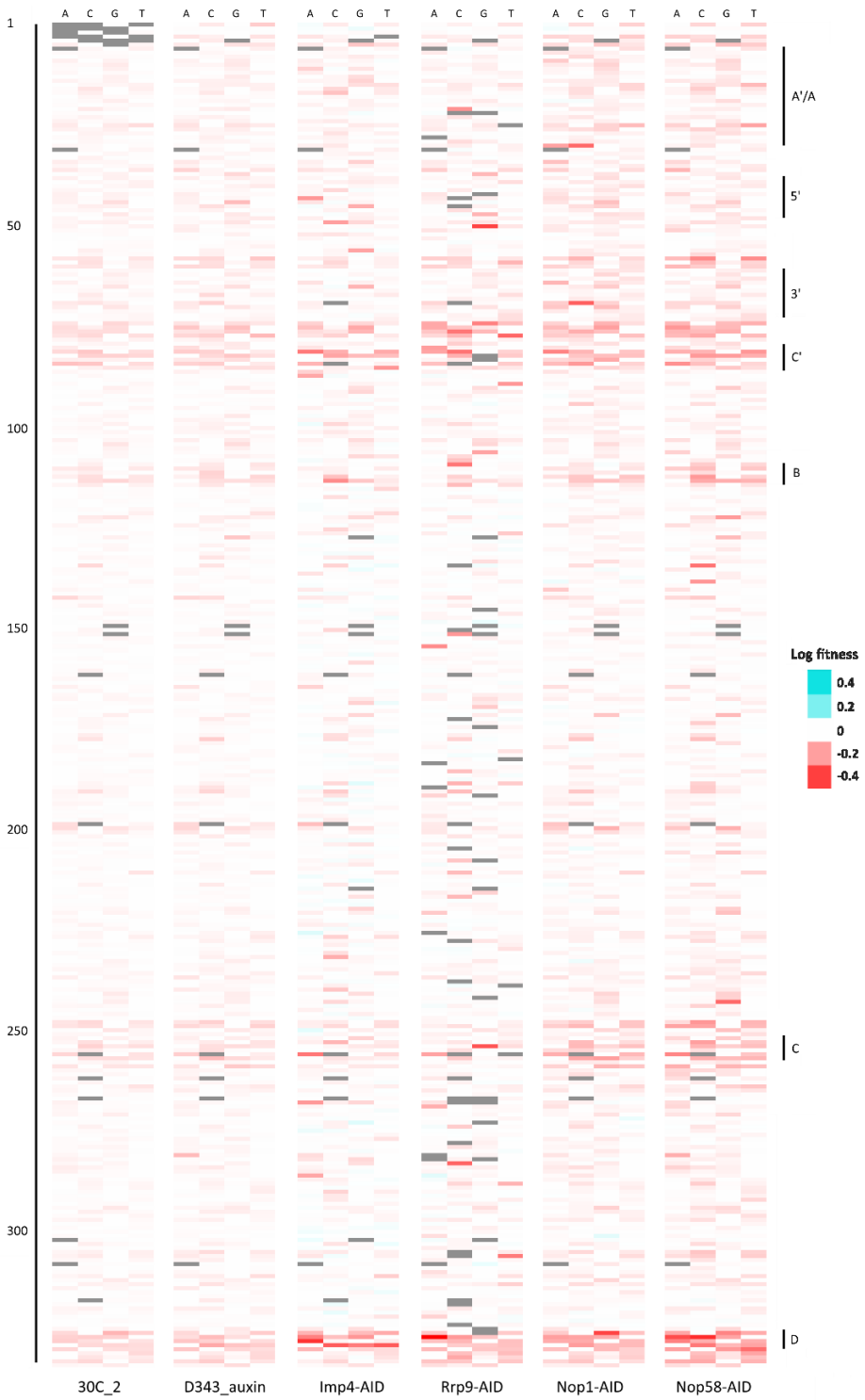


Figure 29, Log fitness values for each substitution in AID strains

Each row is a U3 snoRNA position and each column is one of the four substitutions. Deleterious fitness effects are in red, positive in teal, neutral or wild type in white and missing in grey. Nucleotide positions are indicated on the left, and U3 snoRNA regions indicated on the right. The values are from single mutant genotypes, and double mutant genotypes with a low fitness impact background.

One nucleotide position which does present both heterogeneity and strong intergenic epistasis is position 31. This position is flanking the A box motif, and is four bases away from positions that form a duplex with the pre-rRNA. The log fitness effect for each substitution or single deletion at position 31 is shown for each experiment in (**Figure 30A**). Mutation is tolerated at this robust position in all backgrounds apart from a hypomorphic mutation in Nop1 in which it is significantly deleterious. This is not uniform however, with a mutation to T (U in the snoRNA) having a near 0 log fitness effect, while mutation to A, C or a single deletion leads to a large negative fitness effect. A more subtle version of this pattern can actually be seen in all experiments, with a mutation to T (U) having the highest fitness and C the lowest. This indicates that the cause of the deleterious effect is ubiquitously present, but not tolerated when Nop1 is downregulated. This plot also highlights the slightly higher variation when assigning lower fitness values seen when correlating f_i values (**Figure 26**). The overall pattern is maintained, if not the exact same fitness value.

The cause behind the heterogeneity observed at position 31 may be due to closing an internal loop. A mutation to T (U) maintains the small internal loop in the A'/A box stem, while a mutation to A would cause base pairing with the U at position 11(**Figure 30B**). A mutation to C or a deletion however causes a similar fitness effect. **Figure 30C** shows that due to U3 snoRNA bending upon itself while positioned within the SSU processome, one copy of Nop1 near the box B/C motif is actually positioned very close to position 31. The mutation to T (U) may maintain a similar conformation to the wild type G base, while a mutation to A, C or a deletion changes it. This conformation may be maintained by both the base pairing and Nop1, providing redundancy, but when both Nop1 is missing and the base is changed this can cause synthetic lethality.

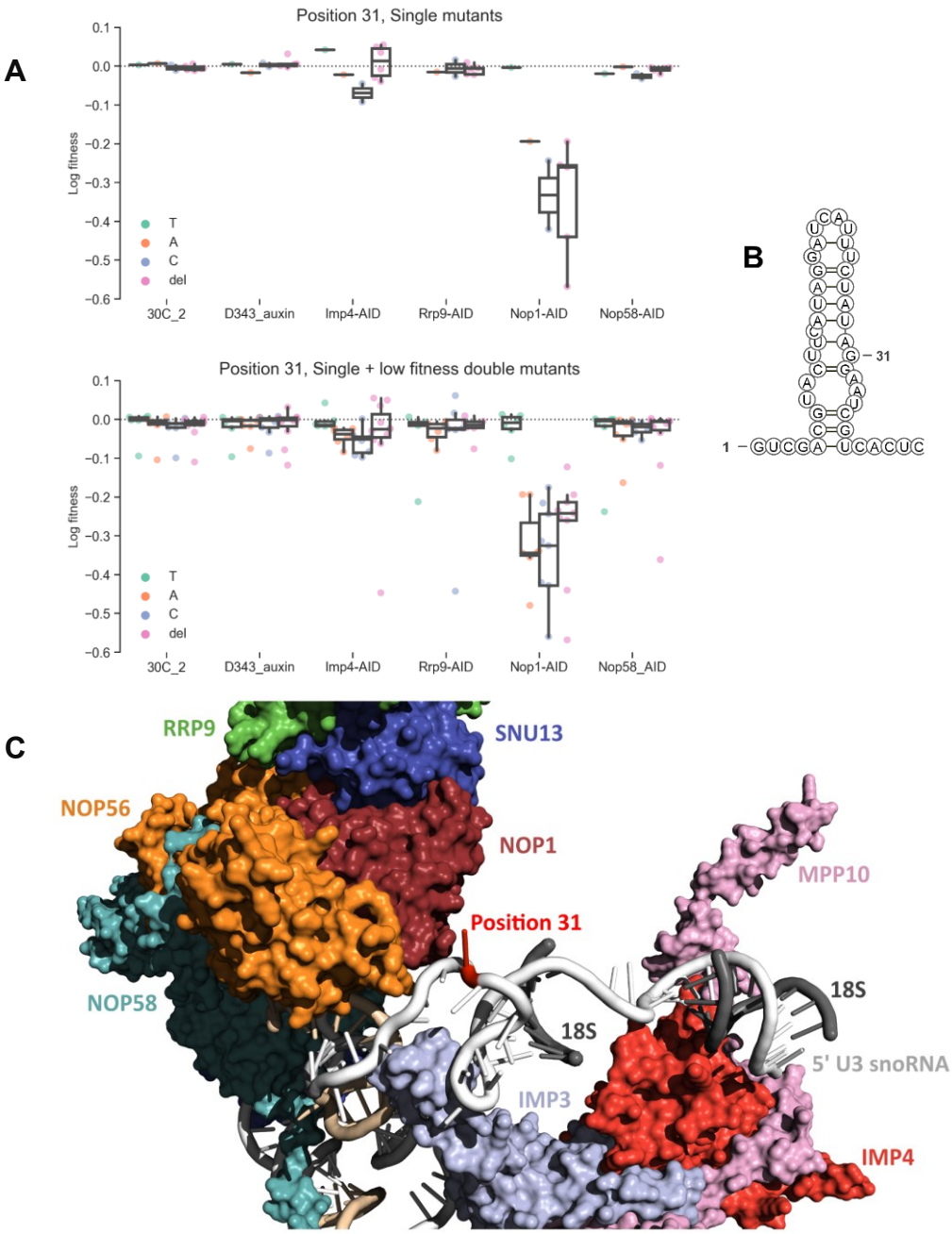


Figure 30, Nucleotide position 31 of U3 snoRNA

A. Box plot showing the difference in fitness effect caused by different mutations (substitutions from G to T, A or C and single deletion) at position 31. The values are from single mutant genotypes and double mutant genotypes with a low fitness impact background. (No single mutants with a mutation to T were present in Rrp9-AID).
B. The secondary structure of the 5' end of U3 snoRNA unbound to pre-rRNA, with bases and nucleotide positions 1 and 31 indicated.
C. A cryo-EM derived cartoon representation of the SSU processome, produced using data from (Barandun, et al., 2017). The following are highlighted: The 5' end of U3 snoRNA (white), the regions of 18S (black) that base pair with U3 snoRNA, Nop1 (rouge), Imp3 (lilac), Imp4 (red), Rrp9 (green), Mpp10 (pink), Snu13 (blue), Nop56 (orange), Nop58 (teal), U3 snoRNA nucleotide position 31 (red).

Nop1 functions as a methyl-transferase in box C/D snoRNAs that are not U3 snoRNA, with 2'-O-methylation of pre-rRNA being impaired when Nop1 is depleted (Tollervey, et al., 1991). A hypomorphic mutation in Nop1 therefore could lead to a reduction in pre-rRNA 2'-O-methylation. The subsequent alternations to pre-rRNA folding and accessibility (Helm, 2006) could be related to the fitness patterns observed. The conformation assumed with a lower level of pre-rRNA 2'-O-methylation within the SSU processome could exacerbate the fitness reduction caused by mutations at position 31. The same pattern may not be seen at position 31 when Nop58 is downregulated as 2'-O-methylation of pre-rRNA is much less affected when Nop58 is depleted (Lafontaine & Tollervey, 1999). Due to the synergistic way in which modifications work, the reduced stability hypomorphic Nop58 confers may still retain enough modifications to not see a phenotype. The effect at position 31 could also be due to one subunit of Nop1 being a binding platform for Fcf2, Sas10 and Utp24 and the other subunit being a binding platform for Utp11 and Bud21 within the SSU processome. The mutation at position 31 may alter Nop1's conformation within U3 snoRNA and subsequently in the SSU processome, making inaccessible to its binding partners.

4.4 Mutations in all U3 snoRNA genetic backgrounds

By incorporating all genotypes in the library into the analysis, the effect of intragenic epistasis can be appreciated, highlighting the effect multiple genetic backgrounds have in shaping the fitness landscape. One way I look at this is by calculating p_i, the aggregate effect of mutations at a position in U3 snoRNA in all genetic backgrounds.

4.4.1 Epistasis with a fitness landscape

The presence of intragenic epistasis is evident when correlating mutations in a wild type background with mutations in all genetic backgrounds (**Figure 31**). The overall relationship is non-linear, with many positions at 0 log fitness in f_i being deleterious in p_i. These positions only exhibit negative fitness effects in the context of mutated genetic backgrounds. The highest density of positions resides at near 0 log fitness, a pattern conserved across all experiments. This again illustrates the robustness of U3 snoRNA in regions such as stem IV, V and V. While the AID strain experiments have an overall negative shift the pattern is still maintained (**Figure 31C-F**). The overall shift seems comparable for all experiments with gene specific differences. The small number of positions that deviate from this pattern are deleterious even in a wild type genetic background, and thus display a more linear relationship. These are positions within the box motif or structurally important stems. The relationship is not completely linear as p_i is normalised to the log fitness of variants that had no mutation at position i.

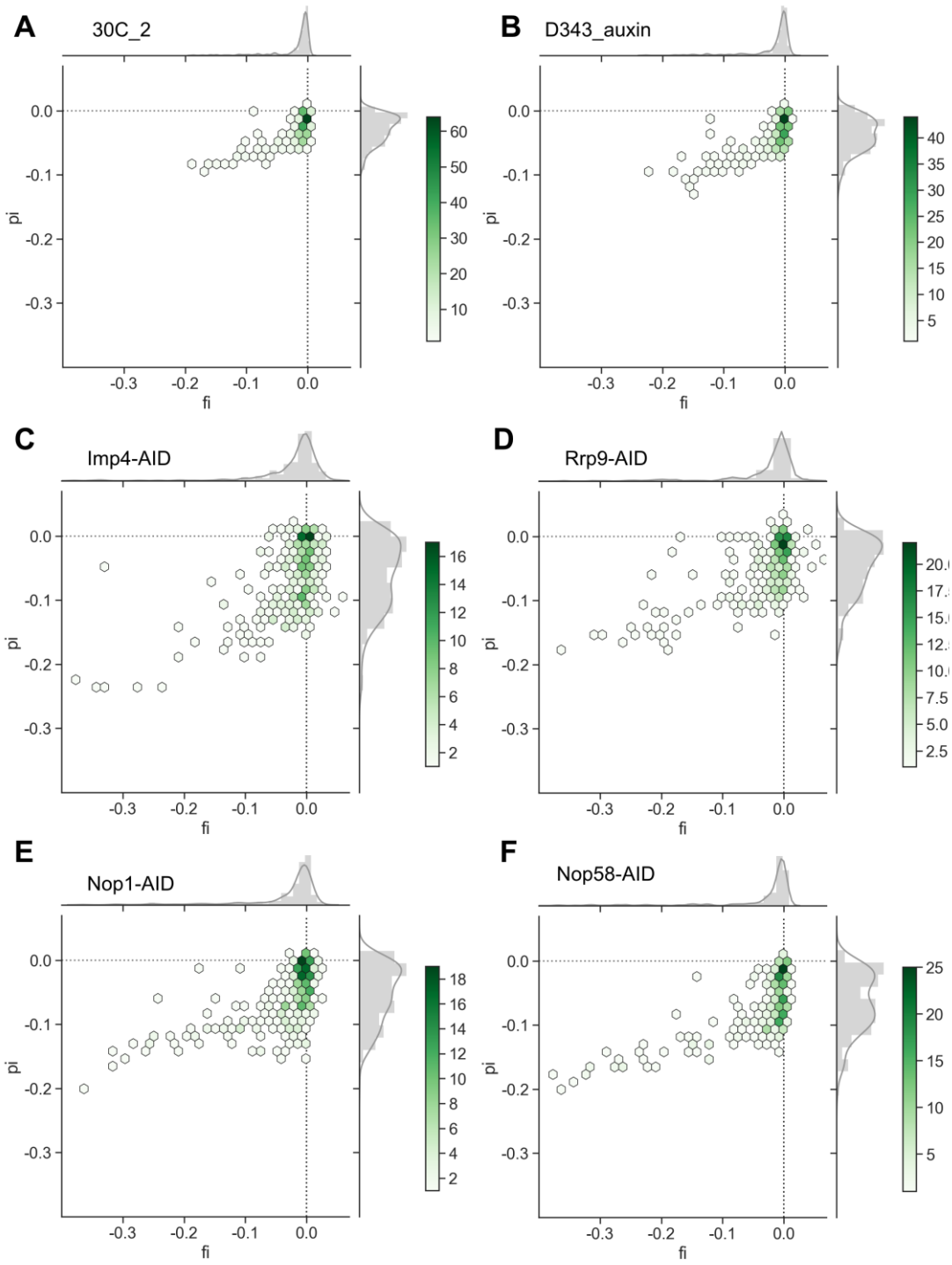


Figure 31, Correlating f_i and p_i values from AID strains

Density hexbin plots of f_i values plotted against p_i values for each experiment, with density plots on the axes. A scale bar for each plot indicates the number of values that constitute each hexagon, with white indicating one value and a darker shade of green indicating a higher value.

When grouping genotypes by mutation number the aggregate log fitness effect decreases with increasing number of mutations, with the effect of 1-6 and more than 6 mutations shown in **Figure 32**. This is pattern is reflected in all experiments. To test whether this is simply the additive effect of single mutations, I also plotted the aggregate fitness effects of single mutants that constitute each of the different multiple mutant genotypes to serve as a control. If there is no intragenic epistasis, the measured effect of multiple mutants should be identical to the aggregation of single mutants. The measured fitness however is lower in all experiments, exposing the presence of intragenic epistasis and the importance of the genetic background in which a focal mutation resides. This pattern appears to disappear when grouping genotypes with more than six mutations. This however is an artefact caused by a saturation effect, where after a certain margin U3 is non-functional, and additional mutations have no effect. The aggregation of single mutants used to calculate the 'no epistasis' values therefore overestimates the possible fitness effect. Examples of this are mutations that reside in the box motifs, as further mutations are redundant in impairing U3's function. This highlights the importance of understanding the underlying biology to accurately identify negative or positive epistasis. The pattern of the no epistasis model being lower at >6 mutations appears to be absent from Imp4-AID (**Figure 32C**). This however is caused by the variation in the Imp4-AID experiment causing some single mutants to be assigned relatively positive fitness values. This then results in underestimation when calculating the 'no epistasis' values. Despite the median fitness decreasing as a function of increasing mutation number in all experiments, there are still many genotypes that remain near 0 log fitness levels. These genotypes consist of mutations located in robust positions.

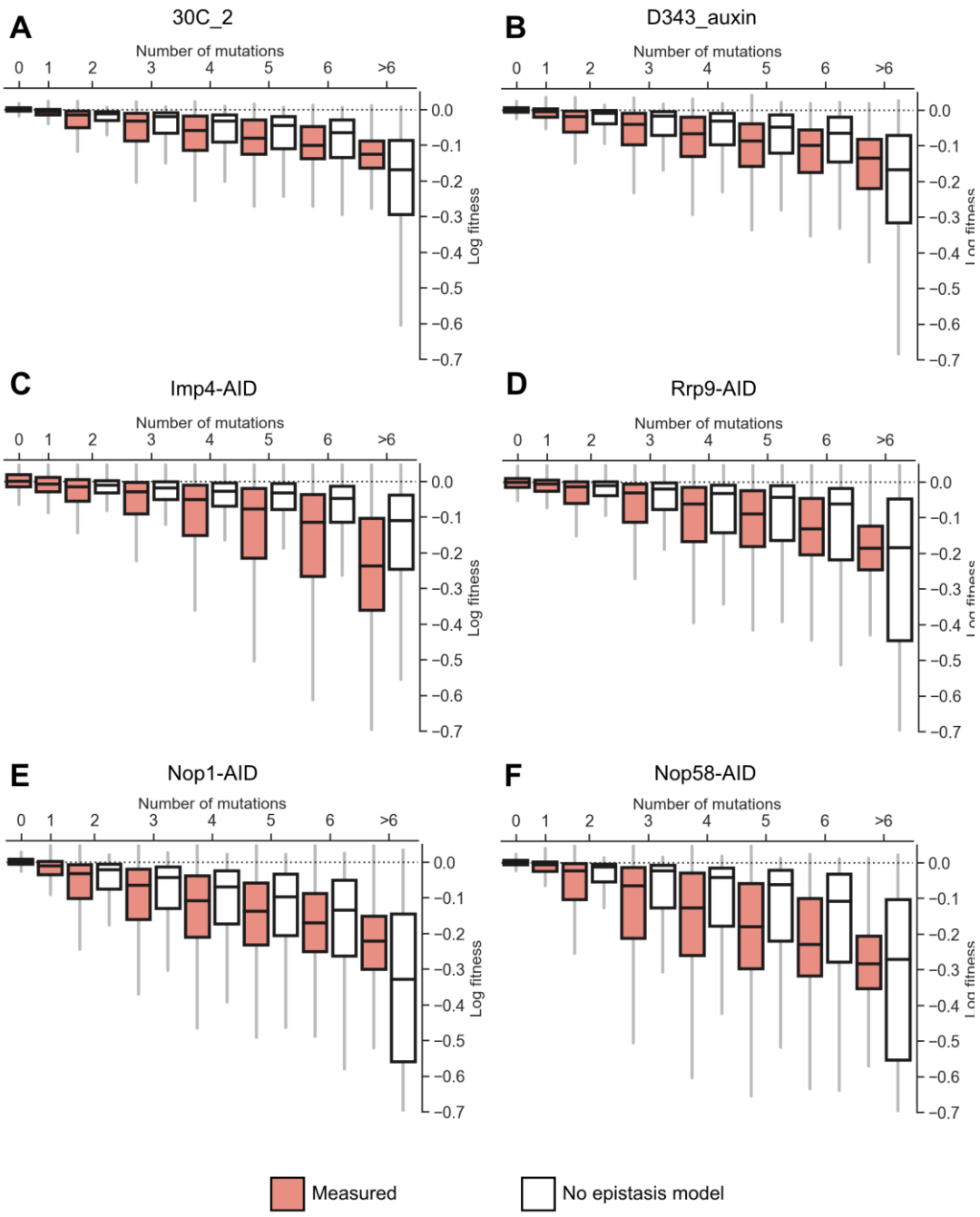


Figure 32, Comparing measured fitness from AID strains to a model with no epistasis

Box plots showing for different numbers of mutations the mean measured log fitness effects (orange) compared to aggregation of the equivalent single mutant genotypes to represent a model with no epistasis (white). The boxes show the median, inter-quartile range, and upper and lower extremes.

To look more closely at the relationship between log fitness and mutation number observed in **Figure 32**, I have plotted the cumulative distribution of log mutants, grouped by the number of mutations present (**Figure 33**). As already observed the proportion of genotypes with a lower log fitness value increases as a function of the number of mutations. This effect is amplified in the AID strains (**Figure 33 C-F**), further diverging from D343_auxin as the number of mutations increases. This was observed when calculating f_i values, with the majority of positions being unchanged when only analysing single mutants (**Figure 27**). This is indicative of a concave fitness landscape, whereby movement in sequence space away from a peak of high fitness progresses down a slope of increasing steepness. When essential proteins are downregulated it appears the slope becomes steeper.

Downregulation of Nop58 results in a larger proportion of genotypes becoming deleterious as mutation number increases (**Figure 33F**). This is perhaps reflective of U3 snoRNA being destabilised in the absence of Nop58, thus having a lower threshold to destabilisation. Imp4 being at hypomorphic levels results in the most linear relationship (**Figure 33C**), as observed in **Figure 32C**, presenting a more intermediate level of destabilisation. Rrp9 (**Figure 33D**) and Nop1 (**Figure 33E**) downregulation leads to a smaller number of genotypes becoming detrimental, indicating specific positions are effected more than overall destabilisation.

To now ascribe more specific molecular mechanisms to the observed patterns, I will now look at which positions and genotypes are effected.

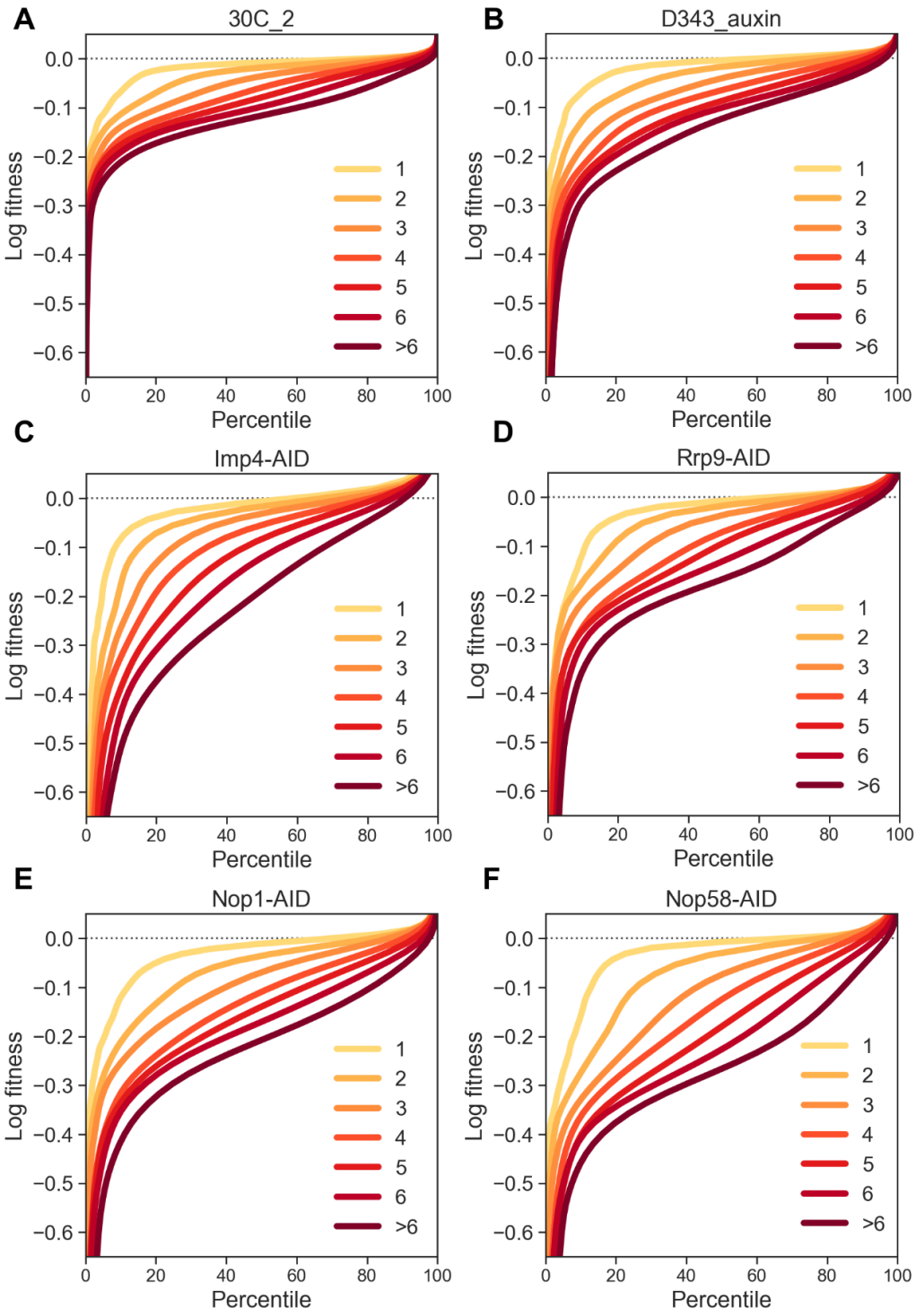


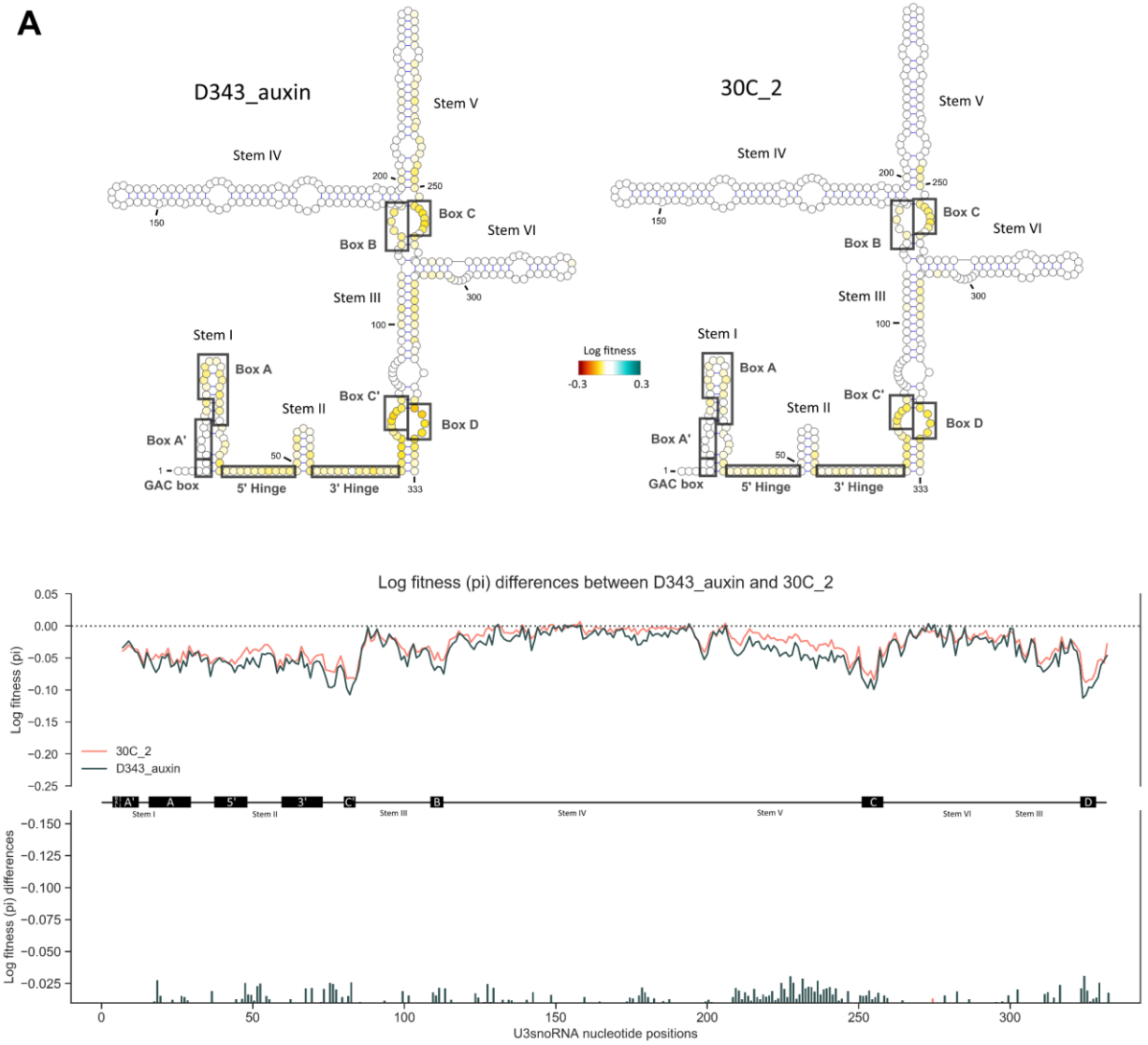
Figure 33, Cumulative distribution plots of log fitness effects from AID strains

Plots for each experiment, showing the cumulative distribution of fitness effects for genotypes, grouped by the number of mutations present.

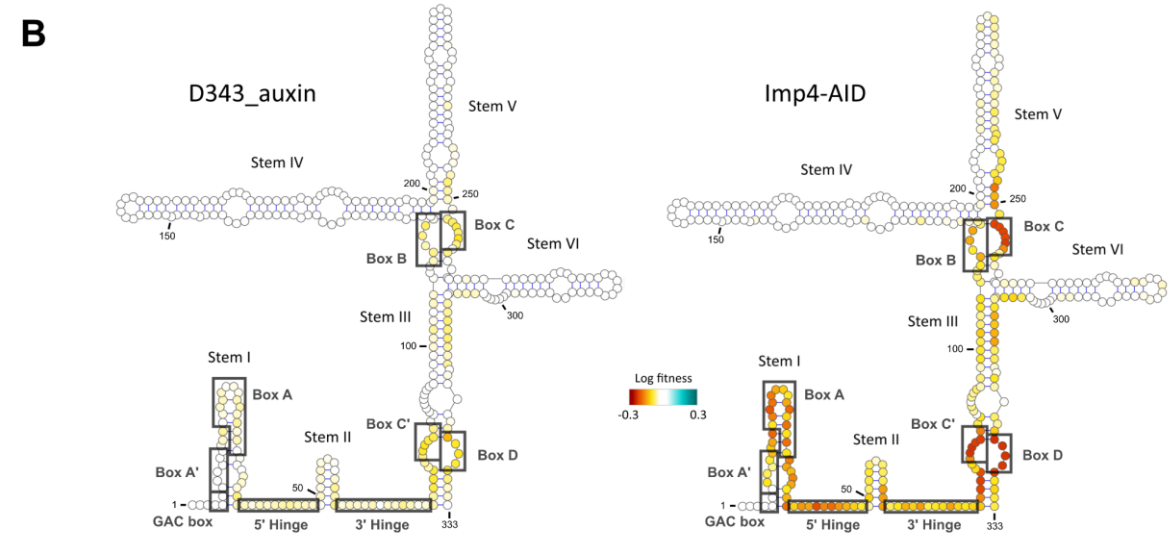
4.4.2 Positional impact of gene-gene interactions in all genetic backgrounds

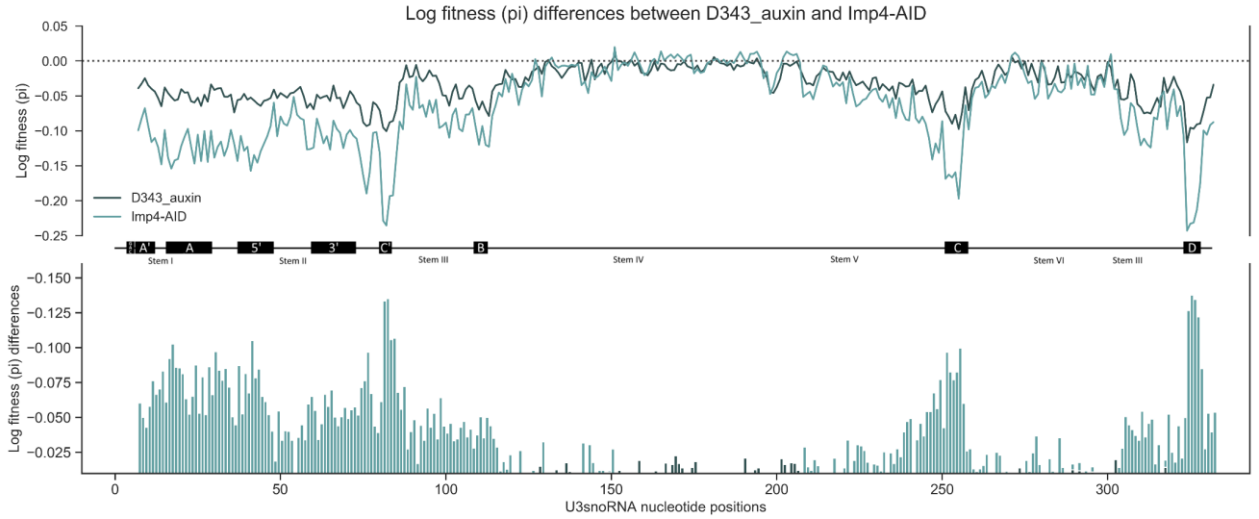
From looking at the overall trends in the data both intragenic and intergenic epistasis are prevalent, with the fitness landscape being altered simultaneously in *cis* by increasing mutation number and in *trans* by hypomorphic mutations in the proteins of interest. To look from a more functional perspective I compared the p_i values of different experiments, calculated as the effect of mutations at a focal position in all genetic backgrounds.

In **Figure 34** I have plotted the p_i values from D343_auxin and the different AID strain experiments using the same genotypes, both on the secondary structure of U3 snoRNA and as a line chart. I have also plotted the difference between the two as a bar chart. First looking at the pattern in the background of wild type protein expression (**Figure 34A**), the inclusion of multiple mutants causes many more positions to show large fitness effects. This is an example of threshold robustness, whereby once an inherent buffer against genetic perturbations is compromised mutations begin to exhibit negative epistasis (Bershtein, et al., 2006). This tolerance to mutation under a certain margin can take forms such as of redundancy afforded by duplicated genes or excessive protein stability. The positions affected by the inclusion of multiple mutants in U3 snoRNA match functional expectations already characterised in U3 snoRNA (Samarsky & Fournier, 1998), located in stem I, II and III, and in the box A'/A motif and two hinge regions that form duplexes with the pre-rRNA. These stems appear to tolerate single changes, but perhaps no longer provide an appropriate structure when numerous changes occur. The fitness effects observed in the A'/A motif and hinge region are due to multiple mismatches being necessary to ablate duplex formation. Despite the increased number of mutations many regions of U3 snoRNA remain robust, with stems IV, V and VI

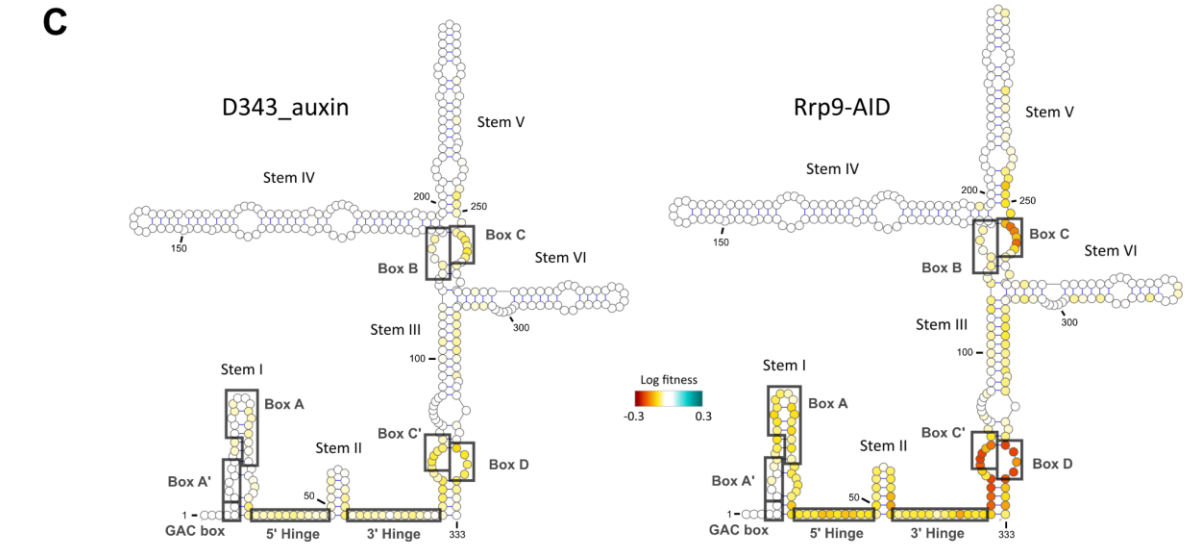

A
D343_auxin
30C_2
Stem V
Stem IV
Stem VI
Box C
Box B
Stem III
Stem I
Box A
Box C'
Stem II
Box D
Box A'
GAC box
5' Hinge
3' Hinge
Log fitness
-0.3
0.3
Log fitness (pi) differences between D343_auxin and 30C_2
Log fitness (pi)
30C_2
D343_auxin
Log fitness (pi) differences
U3snoRNA nucleotide positions

B


D343_auxin
Imp4-AID
Stem V
Stem IV
Stem VI
Box C
Box B
Stem III
Stem I
Box A
Box C'
Box D
Stem II
Box A'
GAC box
5' Hinge
3' Hinge
Log fitness
-0.3
0.3
150
200
250
300
100
50
1
333

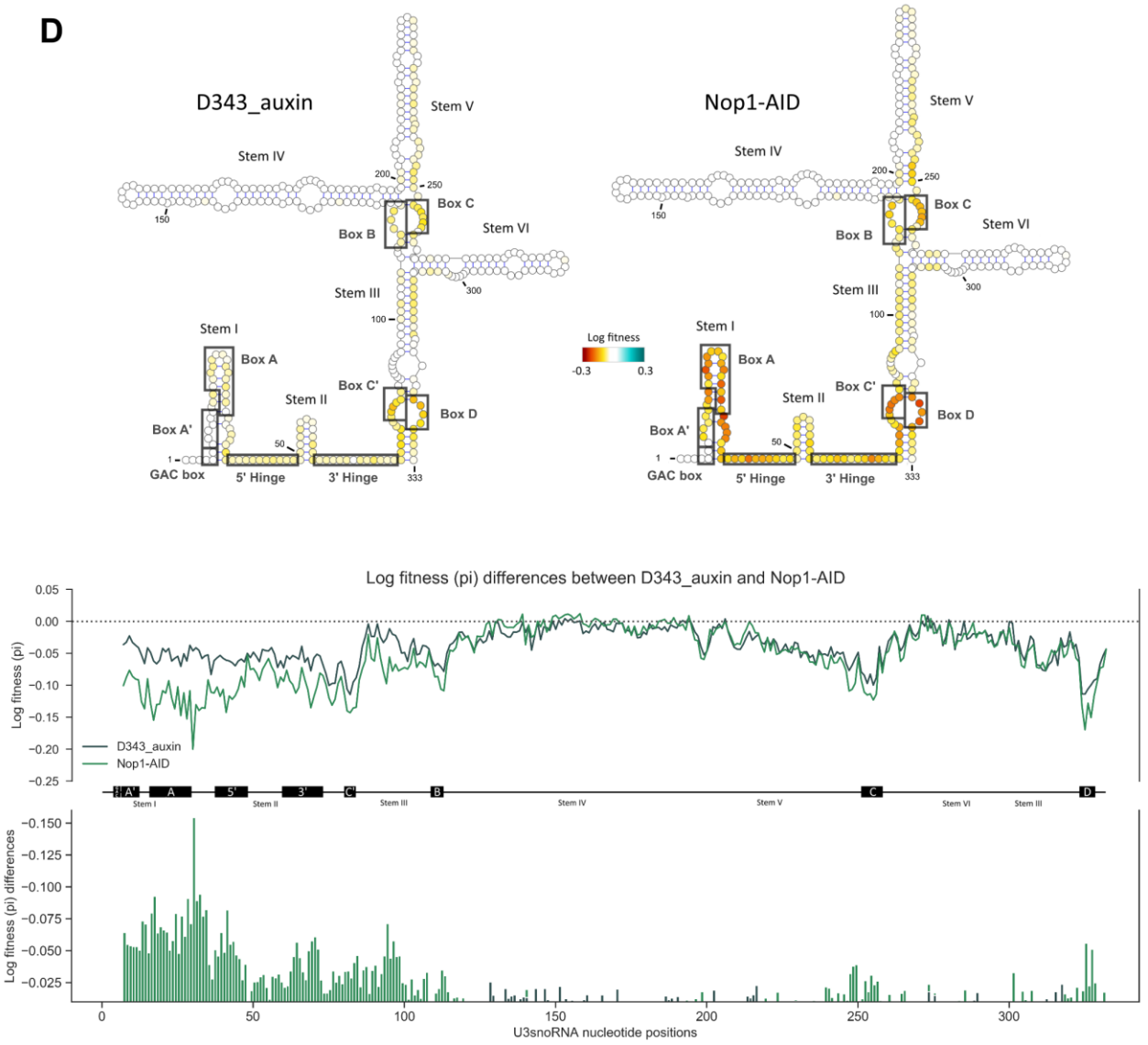

Log fitness (pi) differences between D343_auxin and Imp4-AID
Log fitness (pi)
D343_auxin
Imp4-AID
Log fitness (pi) differences
U3snoRNA nucleotide positions

C


D343_auxin
Rrp9-AID
Stem V
Stem IV
Box C
Box B
Stem VI
Stem III
Stem I
Box A
Box C'
Stem II
Box D
Box A'
GAC box
5' Hinge
3' Hinge
Log fitness
-0.3
0.3
150
200
250
300
100
50
1
333


Log fitness (pi) differences between D343_auxin and Rrp9-AID
Log fitness (pi)
D343_auxin
Rrp9-AID
Log fitness (pi) differences
U3snoRNA nucleotide positions

D


D343_auxin
Nop1-AID
Stem V
Stem IV
Box C
Box B
Stem VI
Stem III
Stem I
Box A
Box C'
Stem II
Box D
Box A'
GAC box
5' Hinge
3' Hinge
Log fitness
-0.3
0.3
Log fitness (pi) differences between D343_auxin and Nop1-AID
Log fitness (pi)
D343_auxin
Nop1-AID
Log fitness (pi) differences
U3snoRNA nucleotide positions

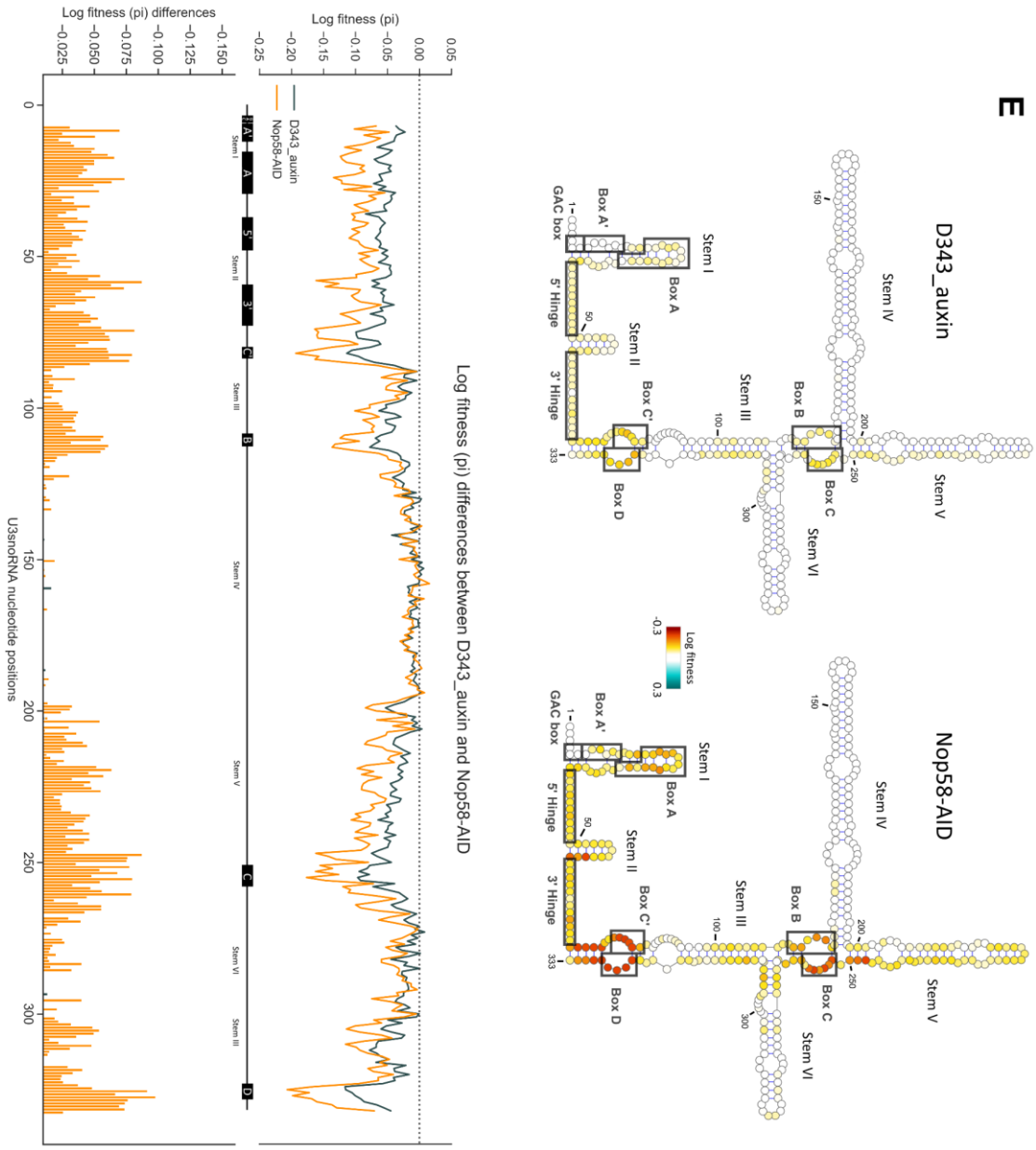


Figure 34, Comparing p_i values from AID strains

U3 snoRNA secondary structure with p_i values plotted on a colour scale (indicated by colour bar), from negative fitness effects (red) to positive fitness effects (blue), with white indicating neutral effects. The p_i values from D343_auxin and the protein of interest strain are calculated with the same genotypes, with the same p_i values from both being plotted on the secondary structure of U3 snoRNA, in a line chart, and the difference between the p_i values for each position is plotted in bar chart. The characterised regions of U3 snoRNA (A box, A' box, hinge regions, C box, B box, C' box, D box) are indicated in a schematic between the line and bar chart. U3 snoRNA nucleotide positions are indicated in 50 nucleotide intervals on the bar chart and on the U3 snoRNA secondary structure.

being very tolerant to mutation. The genetic interactions are not exclusively negative, with positive epistasis existing between base pairs, where a compensatory mutation restores base pairing (Puchta, et al., 2016). This is most strong the terminal stem at the base of the box C'/D motif, with the base pairing being essential both to U3 snoRNA stability and maintenance of the k-turn motif. The fitness landscape of 30C_2 and D343_auxin appear relatively very similar, with the slight fitness decrease previously noted also being represented when calculating p_i values. The log fitness decrease in stem V could be an auxin-dependent effect, or a consequence of a higher sequencing depth as discussed earlier.

The majority of intergenic effects observed in the AID strains reside in positions already sensitive to mutation (**Figure 34B-E**). These are not exclusive to regions that directly interact with the proteins, exposing the cascade of effects that can occur when perturbing one part of a complex or a protein with multiple roles. The box motifs and hinge regions are all affected, though to different extents, when downregulating any of the proteins. This might represent the cause of the overall negative log fitness shift observed, attributed to U3's general function being perturbed when one essential component is disrupted. Under the more stringent selection pressures of a downregulated protein it appears the robustness threshold is compromised more easily.

The robustness threshold being lowered is most evident with a hypomorphic mutation in Nop58 (**Figure 34E**). Nop58 is known to be essential to box C/D snoRNA stability (Lafontaine & Tollervey, 1999), and the overall robustness of U3 snoRNA appears much more easily exhausted. When calculating p_i the intergenic epistasis is the most evenly disributed across all regions compared to other genes (**Figure 35**). This reflects the large proportion of genotypes being assigned a low fitness effect (**Figure 25**)(**Figure 33**). With the stability

Figure 35, Heatmap for p_i values from AID strains in different U3 snoRNA regions

Each row is a U3 snoRNA region (box motifs, hinge regions, GC base pairs and remaining positions), and each column is values from the D343_experiment and protein of interest experiment calculated using the same genotypes. The value is the mean of p_i values in that region, with the colour scale ranging from red for negative log fitness effects, blue for positive log fitness effects and yellow for neutral.

of U3 snoRNA compromised, it appears any many more positions become vulnerable. The most effected positions are same G-C base pairs identified with f_i values, in stem I, base of stem II, base of stem V, base of stem VI and the terminal stem. In addition base pairs in stem III and stem IV are now sensitive, and unique to Nop58, the whole of stem V including the hairpin loop no longer tolerates mutation (**Figure 34E**). With other proteins there is heterogeneity between which box motifs are most effected, however the log fitness difference is fairly uniform between D343_auxin and Nop58-AID (**Figure 35**).

Comparing unpaired regions, G-C pairs, AU / wobble base pairs and box motif / hinge regions can reveal the properties of U3 snoRNA most affected by gene-gene interactions (**Figure 36**). All base pair types show significant interactions between U3 snoRNA and Nop58. G-C pairs exhibit the most significant divergence from D343_auxin (p-value = $9x10^{-7}$), the most significant of all the experiments. This again reinforces the observation that U3 snoRNA is destabilised. This pattern of all base pair types becoming more sensitive to mutation is also seen to a lesser extent with Imp4-AID, with box motif / hinge regions being the most significantly divergent (p-value = $2.5x10^{-15}$). The overall effect is less than seen with Nop58 downregulation. With Nop1-AID and Rrp9-AID the regions most significantly affected are the box motif / hinge regions (p-value = $2.3x10^{-12}$ and $1.1x10^{-10}$ respectively), with other base pair types being only intermediately affected (**Figure 36**). This indicates the overall robustness is compromised to a greater extent when Imp4 or Nop58 are at hypomorphic levels, while interactions are enriched in specific regions for Rrp9 and Nop1. The two 30°C repeats show very little difference between base pair types, while D343_auxin does present a slight difference.

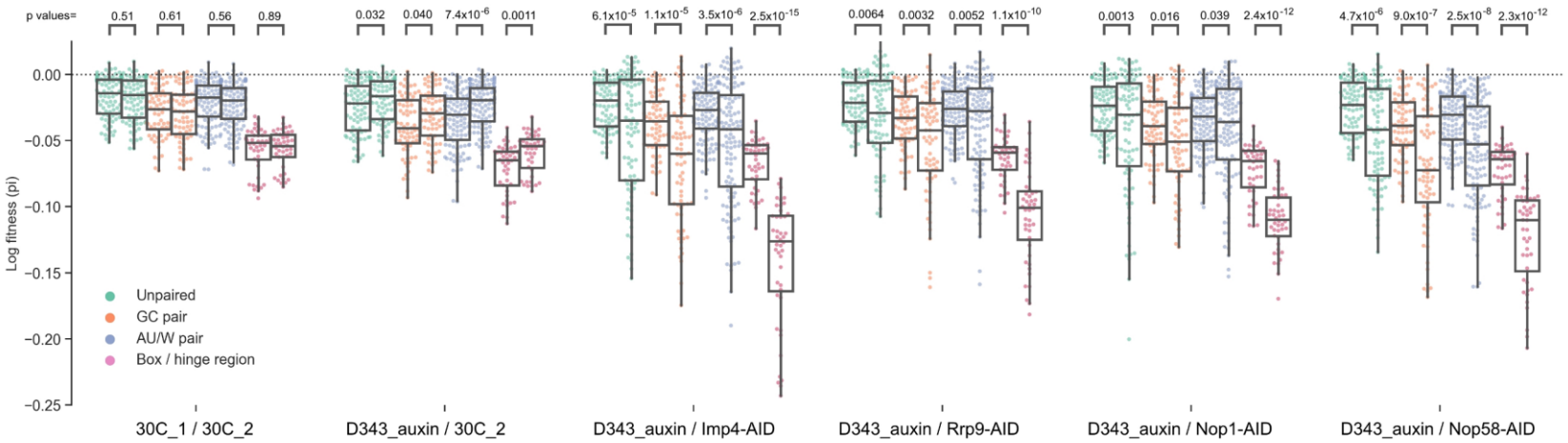


Figure 33, Box plot comparing different position type pi values from AID strains

Comparing the pi values from different experiments in positions that are either unpaired (green), a GC pair (orange), an AU / wobble base pair (blue) or in the box motif / hinge region (red). Student's t-test p-values between experiments are indicated.

This effect is also present when comparing the median log fitness value assigned to the wild type genotypes from each experiment (**Table 4**). The wild type genotypes when Nop58 is downregulated are assigned the highest relative log fitness value (0.141). This could reflect wild type genotypes growing the fastest relative to the other experiments, outcompeting all other genotypes to a higher degree. This effect is not seen with Imp4-AID however. This pattern for Nop58 was observed when plotting the cumulative distribution of log fitness grouped by number of mutation (**Figure 33**), with 72% of genotypes being assigned a low fitness value when Nop58 is downregulated, compared to 65% in D343_auxin

Table 4, Wild type genotype fitness values

Experiment	Number of wild type genotypes that passed filters	Median wild type genotype fitness
30C_2	3039	0.062
D343_auxin	3025	0.082
Imp4-AID	2817	0.103
Rrp9-AID	1751	0.101
Nop1-AID	3050	0.116
Nop58-AID	3056	0.141

When looking at the distribution of p_i values on the secondary structure of U3 snoRNA, hypomorphic mutations in Imp4 and Rrp9 initially appear to cause a similar intergenic epistasis pattern as Nop58 (**Figure 34B-C**), however it is much more region specific. Genotypes with mutations in conserved regions have a large fitness reduction, while robust areas of U3 snoRNA remain more unaffected. Imp4 downregulation affects the hinge regions, stem I, base of stem II, stem III, base of stem V and all box motifs (**Figure 34B**). With Imp4 the fitness decreases relative to D343_auxin are the most dramatic, having the lowest fitness effect for box D, box C, Box C', hinge region and A'A box (**Figure 35**). The distribution of fitness effects is not as ubiquitous as seen with Nop58

however, as robust stems are less effect. While Imp4 is not implicated in U3 snoRNA stability, maintaining a compatible structure may be important. The k-turn constituted of the box C'/D motifs seems most sensitive in the background of a hypomorphic Imp4 mutation. The 5' region of U3 snoRNA does present significant fitness reductions as was expected due to Imp4's location here during the SSU processome (**Figure 23**). The log fitness reduction at the box motifs may represent perturbations caused by general SSU processome disruption, while the large fitness reduction at the A'/A box may be impairing the Mpp10 complexes' function. Imp4 and Imp3 provide a binding platform for Mpp10 to then complete the Mpp10 complex. While Imp4 only plays a small role in the initial unwinding of the box A'/A stem, it plays a major role in the subsequent dissociation (Shah, et al., 2013). It could be expected that mutations that destabilise this stem could compensate for the absence of Imp4, however it appears the general disruption to this region outweighs any positive consequence of destabilisation.

Rrp9 downregulation affects the hinge regions, stem I, base of stem II, base of stem V and all box motifs apart from the box B motif. This is interesting as the box B motif is known to contribute to Rrp9 binding (Venema, et al., 2000 , Granneman, et al., 2002). Effects are mainly enriched in conserved regions; with base pairs outside of these being only intermediately affected (**Figure 36**). Removing the 7bc loop from Rrp9 ablates its ability to recognise U3 snoRNA, leading to lethality (Zhang L, 2013). A hypomorphic mutation may cause a similar phenotype, subsequently making accurately assembled U3 snoRNPs more important due to the low abundance of Rrp9 which is required for the assembly of other SSU processome components. This results in all the conserved regions showing more detrimental fitness effects when mutated in the background of hypomorphic Rrp9 expression (**Figure 35**).

The distribution of fitness effects when Nop1 is downregulated cluster around the 5' end of U3 snoRNA (**Figure 34C**), with only the 5' side of the terminal stem and stem III affected. This gives plausibility to my previous idea that the disruption of Nop1's function in modifying the pre-rRNA may have an effect, as the 5' end of U3 snoRNA is the region which forms duplexes with the pre-rRNA. The changes in conformation caused by the disruption pre-rRNA modification may necessitate more perfect complementarity as the pre-rRNA is less accessible. The significant fitness decrease caused by mutations in position 31 noted when looking at single mutations (**Figure 30**) is preserved when looking at multiple mutants. Nop58 is important to box C/D snoRNA stability, but it is pre-rRNA processing that is affected much more significantly rather than pre-rRNA modifications when depleted (Lafontaine & Tollervey, 1999). The fact that box motifs are much less affected with Nop1 hypomorphy relative to the other proteins tested (**Figure 35**) may indicate Nop1 is less vital to a functional U3 snoRNP, or that Nop1 can dissociate and associate between otherwise complete U3 snoRNPs to complete its role. Nop1 could be sequestered by U3 snoRNPs, leaving little for other box C/D snoRNAs, resulting in pre-rRNA modifications being heavily affected. This could explain the regions of U3 snoRNA that form duplexes with the pre-rRNA being most sensitive to mutation, but other function regions of U3 snoRNA are less affected.

The disparity between a proteins location in the SSU processome and the positions of U3 snoRNA affected by intergenic epistasis is highlighted in **Figure 37**. For Rrp9 (**Figure 37B**) and Nop1 (**Figure 37C**) the conserved regions affected are not proximal to their location. The pattern seen with Imp4 (**Figure 37A**) may be a combination of both protein binding sites affected and functional consequences. Interestingly with Nop58 (**Figure 37D**), an internal loop in stem III which shows little intergenic epistasis is very proximal to Nop58's location in the SSU processome . Disruption of this loop may in some

way be compensating for the low levels of Nop58, as Nop58 is not present to make the conformational changes. It may also be that proteins other than Nop58, which are unaffected, remodel this loop.

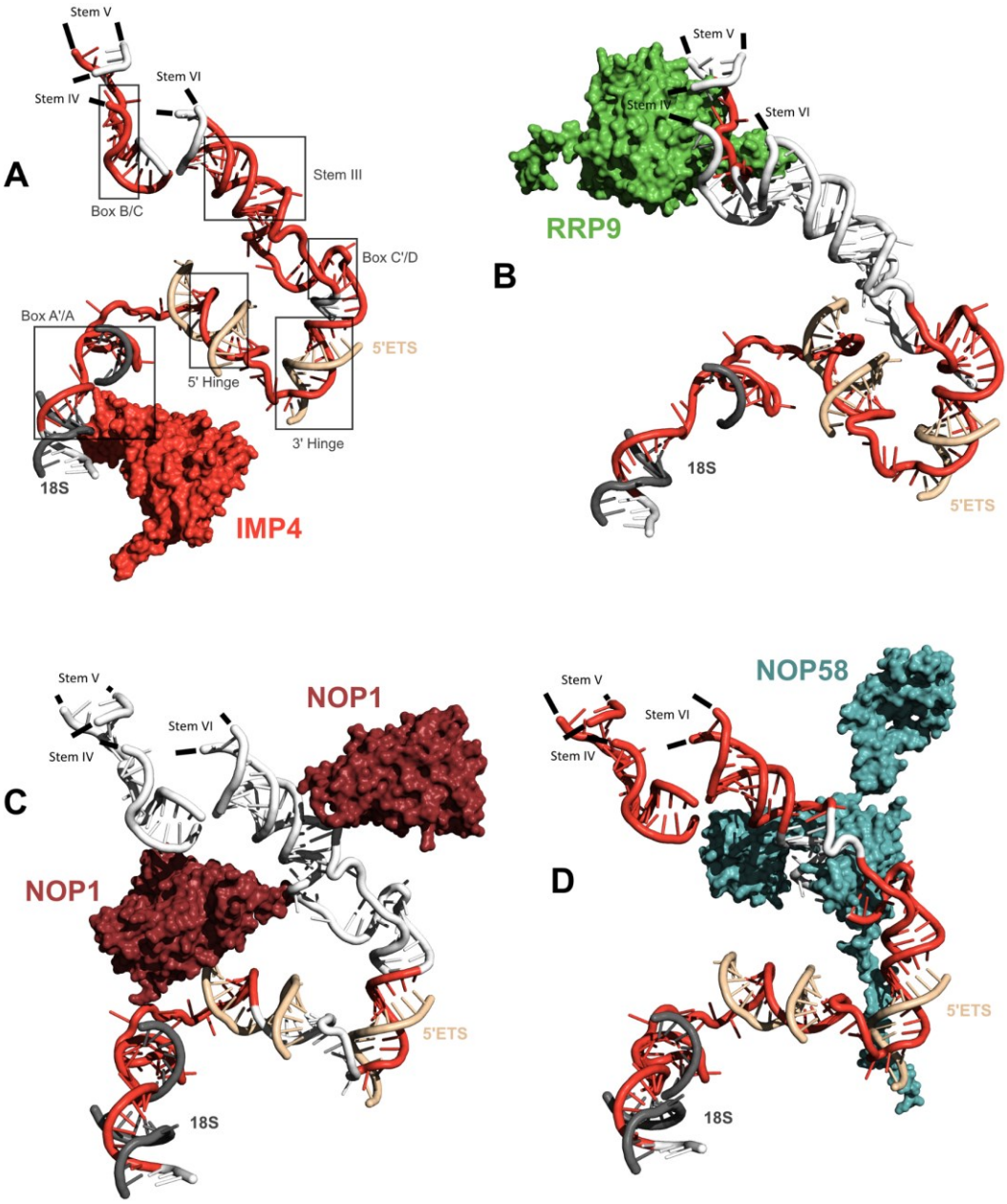


Figure 37, Relationship between protein position within the SSU processome and intergenic epistasis (Cryo-EM)

A cryo-EM derived cartoon representation U3 snoRNA with the protein of interest, produced using SSU processome data from (Barandun, et al., 2017). U3 snoRNA is in white, with Imp4 (**A**) (red), Rrp9 (**B**) (green), Nop1 (**C**) (dark red) and Nop58 (**D**) (teal). For each protein the areas of U3 snoRNA most effected by intergenic epistasis are highlighted in red. This is highlighting regions, rather than the specific nucleotides affected. For Imp4 regions of U3 snoRNA are labelled with boxes, which applies to all four. Stem IV, stem V and stem VI were not captured in this cryo-EM, but where they would be is indicated.

4.5 Discussion

Fitness landscapes give insight into the relationship between genotype and phenotype. Looking at a large breadth of genotypes and the fitness implications can answer key questions in genetics, exposing how mutations manifest to produce a phenotype. While it is now feasible to systematically survey intragenic interactions within a molecule, intergenic interactions pose a bigger obstacle. Using a high-throughput fitness assay I have derived the fitness landscape of U3 snoRNA in five different genetic backgrounds. This has exposed gene-gene interactions between snR17a (U3 snoRNA) and the genes of U3 snoRNA interacting proteins, enabling me to ascribe molecular mechanisms. The level of hypomorphy I have used results in a large reduction in protein levels, with a western blot for Dhr1 indicating an ~80% reduction (**Figure 17**). This large reduction of essential proteins resulted in strong gene-gene interactions.

As the proteins of interest are involved in both U3 snoRNP and ribosome biogenesis, there are many molecular explanations for the gene-gene interactions observed. I initially expected the differences between the fitness landscapes to represent positions involved in the binding of the protein. This would be in the box motifs, or perhaps mirroring the positions identified in cross linking studies (Granneman, et al., 2009), or flanking a protein's location identified in cryo-EM studies (Barandun, et al., 2017). I instead found the patterns of intergenic epistasis also reflected the function of the protein. This reflects the interconnected nature of these proteins, with a cascade of negative effects resulting from their disruption. This is reflected in (**Figure 37**), where the location of the protein within the SSU processome does not correlate with the location of negative fitness effects within U3 snoRNA. Even in these stressful conditions, large regions of U3 snoRNA such as stem IV and stem VI maintain genetic robustness. These regions tolerate multiple mutations with no

intergenic epistasis, showing no fitness difference to the wild-type genotypes. Another interesting observation is certain stems tolerate mutations on one side but not the other. The base of stem II at positions 49-51 are relatively tolerant to mutation, while the complimentary positions 59-61 are not. Similarly at the base of stem V positions 199-201 are relatively robust, while positions 248-250 are not. This pattern is seen in 30C_2, indicating the cause behind this is exacerbated by gene-gene interactions. the. With hypomorphic levels of Imp4 for example, positions 199-201 in stem V are completely unaffected, while positions 248-250 present strong effects. In the terminal stem the 5' side is affected by gene-gene interactions, while for Nop58 the entire stem is. If the negative consequence of perturbing these stems is due to stability, this may represent that one side of the stem is stabilised by a protein interaction. It could also mean that one side of the stem is a protein recognition site, and mutations disrupt its recognition.

Under physiological conditions intragenic epistasis was found to be prevalent in fitness landscape U3 snoRNA, and I have found intergenic epistasis to also be abundant. For the most part gene-gene interactions are enriched in the same positions as intragenic epistasis, highlighting their functional importance. With increasing mutation number the magnitude of effects deviates further from the control, exposing the combinational effect of intra and intergenic epistasis, compromising even previously robust positions. It appears that the increased number of mutations makes U3 snoRNA more susceptible to gene-gene interactions.

4.5.1 Intergenic interactions between genes of interest and U3 snoRNA

Nop58 is known to bind to Nop1 and Snu13, which are associated with the box C'/D motif (Barandun, et al., 2017). The pattern of intergenic epistasis when Nop58 is at hypormorphic levels does not solely reflect this interaction. Instead the pattern reflects the reduction in stability of box C/D snoRNAs that is observed when Nop58 is depleted (Lafontaine & Tollervey, 1999). This is shown in a position independent manner as the log fitness of genotypes decreases as the number of mutations per genotype increases. Mutations in conserved pre-rRNA and protein binding regions as well structurally important stem regions become more sensitive to mutation, with a relatively uniform scale of effects observed.

While many stem structures are affected when the other proteins are downregulated (in particular Imp4), indicating an increased importance of correct secondary structure, the stem structures affected is more severe and widespread when Nop58 is at hypomorphic levels. Even when calculating f_i with single mutations many structurally integral areas become sensitive to mutation. When calculating p_i the entirety of the previously robust stem V is affected. This is indicative of a destabilised molecule, with a much lower thermodynamic threshold for perturbation. This is also reinforced by the increase of intergenic epistasis as the number of mutations within a genotype increases, showing a somewhat position-independent sensitivity to mutation. The p_i values in unpaired, AU/W pairs and GC pair positions are all the most significantly divergent from D343_auxin than any other experiment.

Hypomorphic levels of Imp4 exhibit the largest positional effects. In particular the box C'/D motif shows the largest fitness deviation from the D343_auxin. The conserved box motif and hinge regions are also the most significantly

different from D343_auxin than any other experiment. The secondary structure also seems more important, with all other base pair positions showing significant divergence from D343_auxin. As might be expected with my original hypothesis, the hinge and box A'/A motif where Imp4 is adjacent to within the SSU processome exhibits large intergenic epistasis effects. Even single mutations in positions in the box A'/A and 5' hinge near Imp4's location become sensitive to mutation. I also thought that positive epistatic effects could reside here, as destabilisation of the box A/A' stem could compensate for the reduced activity of Imp4. As Imp4 is required for destabilisation of the A'/A stem to satiate the *in vivo* demand. This does not appear to be the case however, or at least the negative consequences caused by mutation, perhaps obscuring protein recognition sites, takes precedence. It was found that mutations in the hinge regions could rescue a cold sensitive mutant of Dhr1 by destabilising the hinge-ETS duplex (Sardana, et al., 2015). The hinge region however is unstructured, perhaps making the inherent detrimental effect of mutation less than in the structured box A/A' stem.

It may be that the intergenic epistasis pattern seen with Imp4 is reflective of both. That both disrupting Imp4 interaction sites (The box A/A' and 5' hinge), as well as the overall secondary structure of U3 snoRNA (box motifs and stems) is less tolerated with hypomorphic levels of Imp4. The genetic interactions with the conformation of U3 snoRNA may be because Imp4 is missing or only transiently present in the SSU processome, making a conformation suitable for Imp4 binding more important. With the reduced activity of Imp4 destabilising the U3 snoRNA-pre-rRNA duplex, there is greater reliance on the other proteins involved in this role such as Dhr1 and Imp3. This may explain the large scale and breadth of gene-gene interactions seen, as U3 snoRNA is in a suboptimal situation, and thus the robustness threshold for any further perturbation is lowered.

Intergenic epistasis between Nop1 and U3 snoRNA shows the most specific pattern, with the least uniform distribution of positional effects. The structurally important stems are relatively unaffected, and the box C'/D motif shows only a modest increase in sensitivity. The major effects are all clustered in the box A'/A box, and to a lesser extent the hinge regions. This again for the most part does not appear to reflect the location of Nop1 in the U3 snoRNP, which is found in two copies at both k-turn motifs along with Snu13. Instead, this pattern may be a result of Nop1's role as a methyltransferase in other box C/D snoRNAs. Nop1 is not essential for the accumulation of U3 snoRNA, but is for other snoRNAs that carry out methylation, such as U14, U18, U24 and snR190 (Tollervey, et al., 1991 , Lafontaine & Tollervey, 2000). This leads to an accumulation of unmethylated and unprocessed pre-rRNA when Nop1 is depleted. The hypomorphic levels that I create may lead to only partially methylated pre-rRNA, which may have a non-optimal conformation, as methylation is thought to change and reinforce RNA folding by restricting conformational space (Helm, 2006). This suboptimal pre-rRNA folding combined with perturbations in the regions of U3 snoRNA that form duplexes with the pre-rRNA could be synthetic lethal. One position that presents strong intergenic epistasis is C31. This position is in an internal loop in the A'/A stem, with an opposing U base. Substitution to a T (U) is completely tolerated, while a substitution to an A, C or a deletion has a large negative fitness impact. Any genotypes containing those mutations at this position have very low fitness, regardless of the genetic background. From the cryo-EM of the SSU processome before the A_1 cleavage event, U3 snoRNA is in such a conformation that Nop1 is adjacent to position 31. I speculate that this could also be related to the pre-rRNA conformation. Nop1 may help to stabilise or interact with the box A'/A stem, and the conformational changes to stem A'/A caused by mutations at this position disrupt this interaction. This may not be essential in physiological conditions, but due to the combination of conformational changes caused by partial methylation and low levels of Nop1 this becomes deleterious change. Another region that shows a more

intermediate negative fitness effect is the 5’ side of stem III. The other side of the stem III was found to crosslink with Nop1, located near other box C/D snoRNA’s modification guide sequences. The intergenic interactions in this area may be disrupting Nop1 binding. It could also be that both Nop1 and stem III work concomitantly in their structural roles, requiring the perturbation of both to expose redundancy. This could explain why Nop1 depletion does not lead to a reduction of U3 snoRNA levels, as Nop1 has evolved to only provide a supporting role in the maintenance of the secondary structure. This particular region of stem III isn’t affected with Nop58 downregulation but is with Imp4. Nop1 also serves as a binding platform for Fcf2, Sas10, Utp24, Utp11 and Bud21 within the SSU processome. The genetic interactions in the regions of U3 snoRNA that bind the pre-rRNA may be due the increased necessity for Nop1 to be positioned correctly to bind the these proteins.

Intergenic epistasis caused by Rrp9 hypomorphy initially appears to have a similar pattern to Imp4 but is actually more focused in conserved regions. This may reflect an increased pressure for U3 snoRNPs to be correctly assembled, with both the box motifs and hinge regions affected. It may be expected that this simple pattern of deleterious genotypes simply becoming more so is reflective of the stress of an essential protein being depleted, however this pattern is not seen when Nop1 is downregulated. Rrp9 is not thought to bind Snu13 directly (Zhang L, 2013), but rather recognises the conformation that Snu13 creates. It appears that this region surrounding the box B/C motif is no more sensitive to mutation, with the box B region showing no divergence from D343_auxin. It may require looking at specific genotypes to further investigate this region. This however is difficult with the Rrp9-AID experiment, as the inherent variation makes single variants unreliable. Rrp9 was found to extensively crosslink with stem IV, a stem that is very robust to mutation (Granneman, et al., 2009). I expected that the increased selective pressure caused by Rrp9 being at hypomorphic levels may comprise this robustness,

but this does not appear to be the case. It was observed that a ‘weak’ form of the C’ box rendered U3 snoRNP formation reliant on Rrp9 (Knox, et al., 2011). I looked to see if this C’ box sequence existed in the library, as it would interesting to see the effect both with physiological and hypomorphic levels of Rrp9. Unfortunately, a genotype containing this sequence does not exist in the library.

4.5.2 Influence of Intergenic epistasis on the topology of a fitness landscape

The fitness landscape of U3 snoRNA elucidated in physiological conditions appears to indicate a molecule with well adapted regions separated by regions of comprehensive genetic robustness. The topology of this manifests as wide peak in sequence space with steep sides. Any positive epistatic effects are examples of reciprocal sign epistasis, whereby compensatory mutations restore base pairing. This compliments findings that sign epistasis is abundant and restricts potential evolutionary paths, leading to very contingent evolution. The addition of strong intergenic epistasis caused by hypomorphic mutations in protein binding partners appears to make the sides of the peak even steeper. The wide plateau of genetic robustness is somewhat compromised, as previously neutral genotypes are no longer tolerated. This would seem to further restrict the already restricted evolutionary paths for U3 snoRNA.

4.5.3 AID system

The use of the AID system offers an adjustable level of conditional protein depletion. In the context of the completion experiment the system appears to

be very appropriate. The conditional nature of the system allows protein depletion to coincide with the downregulation of genomic U3 snoRNA, avoiding dominant negative genotypes being removed from the starting population. The only caveat is the auxin-dependent growth defect caused by growth in CSM media. The use of a rich media such as YPG would not be a suitable solution as only around 20% nonessential gene deletions cause a measurable growth defect in rich media (Giaever, et al., 2002), potentially causing a significant reduction in the sensitivity of the assay. It has been noted however that this growth inhibition can be avoided with different media compositions such as minimal Kaiser media (Ochoa, 2017). This may present a direction for any future experiments. The fitness landscape of D343_auxin obtained under a 350μM auxin concentration shows minimal differences to 30C_2 with single mutations, but with multiple mutations exhibits more of a deviation. In this way D343_auxin provides an appropriate control for looking at the gene-gene interactions in the AID strains but is hard to rule out auxin-depended effects on the fitness landscapes. Another cause of the differences may stem from D343_auxin being sequenced at a higher depth. This could result in low fitness variants, which are also low copy number in the library, being given a larger starting read count and thus a larger opportunity to present a decreased read count and be assigned a lower log fitness value.

There is also a ubiquitous negative shift observed, likely resulting from the duress caused by an essential protein being depleted. This could also technical artefact caused by the longer time between time points. While the use of divisions when calculating log fitness appears to compensate well, it may not completely cover the difference in time. The time between time points is longer, even with the same amount of generations. The longer time allows non-functional U3 snoRNA genotypes to disappear from the population, and be assigned a lower fitness value.

4.5.4 Potential follow up work

The next work to be completed is to survey the gene-gene interactions between U3 snoRNA and the remaining AID strains I created. These include Nop56, Mpp10, Imp3, Dhr1. It will be interesting to see if the fitness landscapes of proteins that interact with one another will be similar. For example Imp4 is part of the Mpp10 complex with Imp3 and Mpp10. The Emw1-AID strain should also provide a valuable control, as Emw1 is uninvolved in ribosome biogenesis instead involved in cell wall maintenance. As its growth is comparable to the other AID strains, it potentially could provide a more accurate control fitness landscape to quantify intergenic epistasis. It may then be interesting to see how these results compare to a phylogenetic analysis, to see if intergenic epistasis is clustered in conserved regions.

For Snu13 where I was unable to obtain an AID strain, it would also be worth trying other protein downregulation systems such as the DAmP system or putting the degron tag on the N-terminus. Due to the different selection of mating types caused during sporulation, it would be possible to mate two haploid strains of opposite mating types to create diploid strains, which could then be sporulated. It could be interesting to create strains with two AID tagged proteins.

While the current methodology appears to work well, it may be worth making modifications. For example creating more FLAG tagged strains would facilitate accurate quantification of the protein downregulation, and allow uniform protein depletion between experiments. It could also be advantageous to optimise a media composition which completely avoid the auxin-dependent growth defect, ruling out its influence on the results. For experiments that need repeating, such as Imp4 and Rrp9 due to the inherent variation, it may be worth taking these optimisation steps for protein depletion quantification and removal

of auxin-dependent growth defects. The increased growth rate would also make the strains closer to that of the control strain's growth, making comparisons that are more accurate.

Primarily I use the p_i and f_i values to investigate the fitness landscapes of respective experiments. There are additional ways that the data can be analysed. While there are instances of genotypes being covered by multiple variants, the majority of genotypes are only represented once. In order to account for this, the large number of wild type genotypes could be used to estimate noise as a function of coverage. This noise estimate could then be used to assign confidence scores to genotypes. Biological repeats would also significantly help in this situation, to see if certain genotypes maintain fitness effects.

For U3 snoRNA the network of intragenic epistasis was surveyed (Puchta, et al., 2016). This was possible due to the use of the 'big' library whereby combinations mutations could occur throughout the gene. A consequence of the library creation (covered in **Chapter 2**) however is the number of mutations per variant is distributed higher. I instead used the 'small' library whereby mutations can only occur between 100bp windows of U3 snoRNA, which results in a lower average number of mutations per variant. So while the network of intragenic epistasis cannot be surveyed for the entire RNA, it is possible in certain regions. It may for example be interesting to look at the interactions in the hinge regions, a noted hotspot for epistasis, to see how this is altered by intergenic epistasis. Another region is the B/C box and surrounding positions, as it is not only the sequence but the secondary structure of this region that has been identified as important (Granneman, et al., 2002).

Another potentially interesting experiment would be to complete the competition experiment with galactose media, resulting in both wild-type genomic U3 expression and expression from the mutant plasmid library. This was completed in physiological conditions (Puchta, et al., 2016), which found no significant fitness effects. If completed with AID strains however, the hypomorphic levels of the proteins may expose dominant negative genotypes. These could manifest in scenarios whereby U3 snoRNA could still bind the protein, but duplex formation with the pre-rRNA was disrupted, leading to the protein being sequestered on these non-functional U3 snoRNAs.

It was found that a cold sensitive hypomorphic allele of Dhr1 could be rescued by mutating the 5' end of U3 snoRNA, as the duplex disruption compensated for Dhr1's reduced helicase activity (Sardana, et al., 2015). Only a random assortment of U3 snoRNA mutants were tested. By creating this same Dhr1 allele in the D343 strain and completing the competition experiment, the U3 snoRNA mutants that rescue the strain could be systematically surveyed.

5 INVESTIGATING GENE-ENVIRONMENT INTERACTIONS WITHIN THE FITNESS LANDSCAPE OF U3 SNORNA

5.1 Characterising gene-environment interactions

The environment in which mutations occur is a major determinant on their consequence. This results in the environment having a large influence on an organism's evolutionary trajectory. Detrimental mutations in one environment may be neutral or beneficial in another, leading to environmentally driven movement within the fitness landscape (Flynn, et al., 2013). The variability of natural environments can then lead to selection of only conditionally beneficial alleles. Environmental shifts can also facilitate the crossing of low fitness valleys, functioning as a bridge between two high fitness peaks (Steinberg & Ostermeier, 2016). This environmentally driven alteration to the fitness landscape is illustrated in **Figure 38**, providing a route to a global optima.

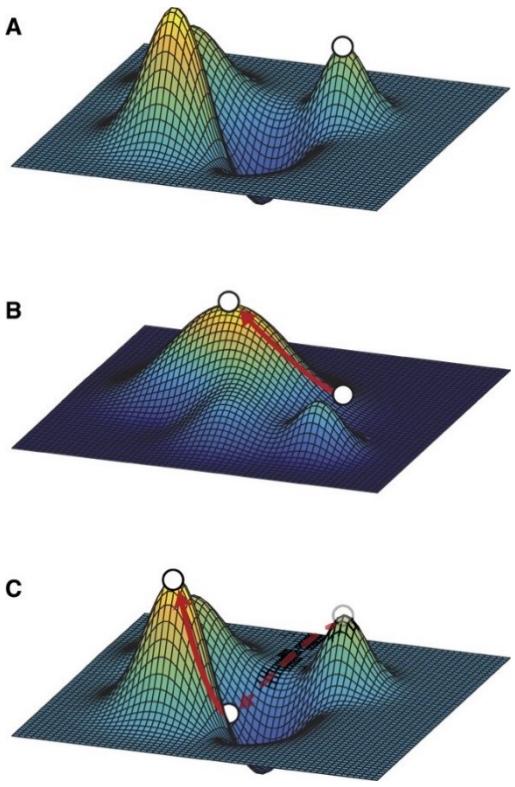


Figure 38, Crossing fitness valleys in a fitness landscape

Model depicting how alternating environments can alter the fitness landscape to facilitate crossing low fitness valleys from (Steinberg & Ostermeier, 2016). The xy plane represents sequence space, the z axis is fitness.

A. The white circle represents a sequence atop a fitness peak, separated from the global optima by a low fitness valley.
B. A environmental shift changes the topology of the fitness landscape, removing the low fitness valley.
C. After shifting back to the previous environment the sequence is now located in the low fitness valley, but is now closer in sequence space to the global optima.

5.1.1 Different temperature environments

Temperature changes are among the most pervasive and influential environmental factors that an organism can face. Along with body size it is the major determinant of biomass production (Enquist, et al., 2003). Temperature can affect both the biophysical and biochemical properties of molecules, necessitating specific mechanisms to mitigate any detrimental effects. Variation beyond a threshold relative to the physiological temperature can elicit an adaptive heat or cold shock response. The exact thresholds vary, with an increase from 50°C to 60°C causing a heat shock response thermophilic bacteria (Daniels, et al., 1984). In yeast with a physiological temperature of 30°C, an increase past 37°C elicits a heat shock response, with further increases in temperature leading to further induction. The response results in a reduction in overall expression levels, and the upregulation of conserved heat shock proteins that contribute to increased thermotolerance of the cell in multiple aspects (Gasch, et al., 2000). These include molecular chaperons to deal with misfolded and aggregated proteins (Nathan, et al., 1997), cell wall restructuring (Martín, et al., 1993) and interruption of the cell cycle (Rowley, et al., 1993). This process is facilitated by different RNA secondary structures and the corresponding folding energies and melting temperatures (Tm) (Wan, et al., 2012). Non-coding RNAs globally have secondary structures with higher thermostability, while RNA thermometers (structures that prevent translation located in 5'-untranslated regions of mRNAs) and ribosomal protein mRNAs have low Tms. This leads to their degradation upon shifting to the heat shock temperature, leading to increased translation of RNA thermometer mRNAs, and repression of ribosomal proteins. At 37°C the yeast heat shock response is adequate to cope with the temperature shift, with no significant protein misfolding or growth defects even when Hsp90, an abundant heat hock chaperone, is depleted (Nathan, et al., 1997). A pre-exposure to 37°C before 50°C results in an increased survival rate, indicating that 37°C is eliciting a heat shock response that allows the cell to acquire themotolerance (Sanchez

& Lindquist, 1990). A temperature of 37°C is commonly used to probe mutant alleles in yeast if no phenotype is seen at 30°C due to the increased stress.

Temperatures environments of 10°C-18°C elicit a cold-shock response in yeast. Similar to heat shock, this results in a change in gene expression (Sahara, et al., 2002). These expression changes are implicated in pre-rRNA processing, ribosome biogenesis and cell wall maintenance. The protein Nsr1 is essential during the cold shock to maintain proper pre-rRNA processing, with levels of 27S, 20S and 7S decreasing at 10°C in a Nrs1 mutant (Kondo, et al., 1992). It is thought Nsr1 could function to destabilise secondary structures, such as those found in snoRNPs, to compensate for the over-stabilisation caused by the low temperature environment (Tronchoni, et al., 2014). Hyperstabilisation of secondary structures is one mechanism behind cold sensitive phenotypes, with cold sensitive mutations in U2 being the result of competing alternative structures (Zavanelli, et al., 1994). This phenotype was suppressed with mutations that disrupted the alternative structure. This can also be advantageous, with alternate secondary structures of mRNAs allowing certain influenza strains to become cryotolerant (Chursov, et al., 2012). Another theory behind why mutants may be cold sensitive but not heat sensitive is due to their involvement in molecular assembly, which is driven by hydrophobic interactions (Strauss & Guthrie, 1991). Bellow 10°C results in a near-freezing response, with differences in gene expression compared to the cold shock (Homma, et al., 2003). This includes the induction of trehalose synthesizing enzymes.

5.1.2 Systematically investigating gene-environment interactions

Deriving a molecule's fitness landscape surveys the relationship between genotype and fitness. For a given environment, this allows both the fitness impact of mutations and as well as the interactions between them (intragenic epistasis) to be quantified. By assaying the fitness landscape of a molecule in a different environment, the interaction of said environment with mutations (gene-environment interactions) and intragenic epistasis (intragenic epistasis-environment interactions) can be exposed.

The fitness landscapes of tRNA-Arg(CUU) in multiple environments revealed abundant gene-environment and intragenic epistasis-environment interactions (Li & Zhang, 2018). The negative epistasis found at the physiological temperature was still present, with the environmental interactions resulting a subsequently predictable pattern. For a ribozyme both environmental interactions and intragenic epistasis worked in conjunction to shape the fitness landscape (Hayden & Wagner, 2012).

To systematically survey gene-environment interactions within U3 snoRNA, the competition experiment was completed in different temperature environments. This provided the fitness landscape of U3 snoRNA in three environments: 30°C, 18°C and 37°C. 30°C represents the optimal growth environment for the D343 strain, 18°C presents a cold stress and 37°C presents a heat stress. The experiment was completed as previously detailed (**Figure 8**), with the only difference being the temperature at which the yeast were cultured in after the first time point. The temperatures were chosen as RNA stability and folding are temperature dependant. By completing the assay at higher temperatures I predicted destabilising mutations will have increased potency, exposing nucleotides vital to maintaining structural integrity. Conversely at low temperatures, destabilising or stabilising mutations in

regions requiring conformational plasticity may have positive or negative fitness affects respectively. The specific base pair conformations of U3 snoRNA are essential to its function as proteins recognise the conserved structural motifs. Different temperature environments can therefore expose which positions are most essential to the maintenance of these structural elements. The temperatures were also chosen due to their physiological relevance, being at the thresholds where yeast incite a response. At 37°C the yeast produces a heat shock response and 18°C produces a cold shock response.

The fitness landscape datasets I will be using include two 30°C repeats from (Puchta, et al., 2016), “Small_1_30C_Glu” and “Small_3_30C_Glu” which I refer to as “30C_1” and “30C_2” respectively. I use the fitness landscape from 30C_2 to compare to other temperature environments as it was constructed from more time points. I will also be using the 37°C data set from that study, “Small_2_37C_Glu”, which I refer to as 37C. I completed the experiment at 18°C, which I refer to as 18C. The different temperature environments result in a difference in growth rate (**Figure 39**). To account for this the log fitness values were calculated with generations as the time component as opposed to hours passed. The OD_{600} measurements taken when the cultures were split were used to calculate the number of divisions that had occurred since the last measurement. The time points taken and the number of divisions at each is shown in **Table 5**. The use of divisions when calculating log fitness for the different variants should account for the large growth defect when the experiment is performed at 18°C.

Table 5, Experimental time points

Experiment Name	Time point	Time (hours)	Divisions
30C_1	1	0	0
	2	36	15
	3	60	25.1
	4	84	35
	5	108	44.9
30C_2	1	0	0
	2	12	4.9
	3	24	9.4
	4	36	14.5
	5	60	24.3
	6	84	34.2
	7	108	44.14
37C	1	0	0
	2	36	14.1
	3	60	23.5
	4	84	32.7
	5	108	41.5
	6	132	50.6
	7	156	59.7
18C	1	0	0
	2	36	6.3
	3	84	13
	4	156	24.8
	5	228	36.2
	6	300	47.6

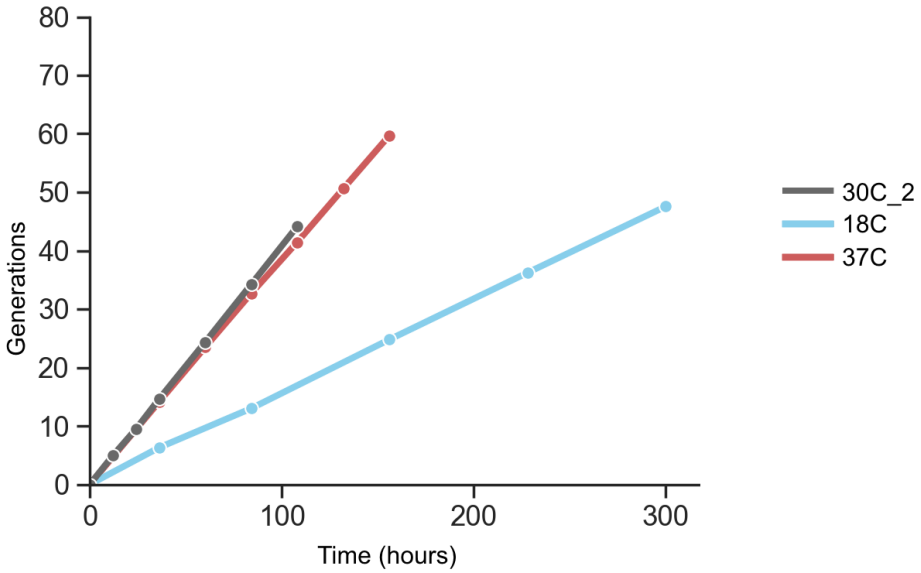


Figure 39, Competition experiment at different temperatures

Line chart illustrating the number of divisions occurring over time (hours) for the 30C_2, 18C and 37C experiments over the course of the respective competition experiment.

5.2 Fitness landscape of U3 snoRNA at different temperature environments

The sequencing results are used to count the number of times a barcode is found in each time point. This gives an approximation of each variants copy number in the population of yeast. Exponential decay curves were then fitted to each barcode's count data, used to measure the logarithm of relative fitness for that genotype. These fitness values are then normalised using the wild type variants present, adjusting neutral mutations to a log fitness value of 0. These results can then be used to construct the fitness landscape of U3 snoRNA in a specific temperature environment, providing the fitness effect for ~24,000 genotypes.

These experiments were performed with the first version of the U3 snoRNA mutant plasmid library, before the addition of variants with mutations in the first six positions of U3 snoRNA. This results in a uniform coverage of mutations across positons 7-333 of U3 snoRNA.

5.2.1 Pattern of fitness effects

The distribution of log fitness effects for a given experiment shows the range over which genotypes lie, and any patterns within the data. For both 30°C repeats the distribution is similar, with the tail of distribution ending at ~-0.3 (**Figure 40A-B**). The largest peak contains genotypes at near 0 log fitness. These genotypes contain mutations in genetically robust areas of U3 snoRNA, such as stem IV, V and VI. The rest of the genotypes contain mutations that effect U3's role in some capacity, usually located in evolutionarily conserved regions. The pattern of distribution is not identical, with 30C_2 exhibiting a

more bimodal distribution. This variation between repeats is due to increased noise when assigning lower fitness values and different genotypes being detected between experiments. A major advantage of this methodology however is the technical noise that causes these differences in distribution can be greatly reduced when aggregating the ~24,000 variants. The percentage of deleterious genotypes is very similar between the two repeats, with 63% of 30C_1 and 65% of 30C_2 being <-0.02 log fitness.

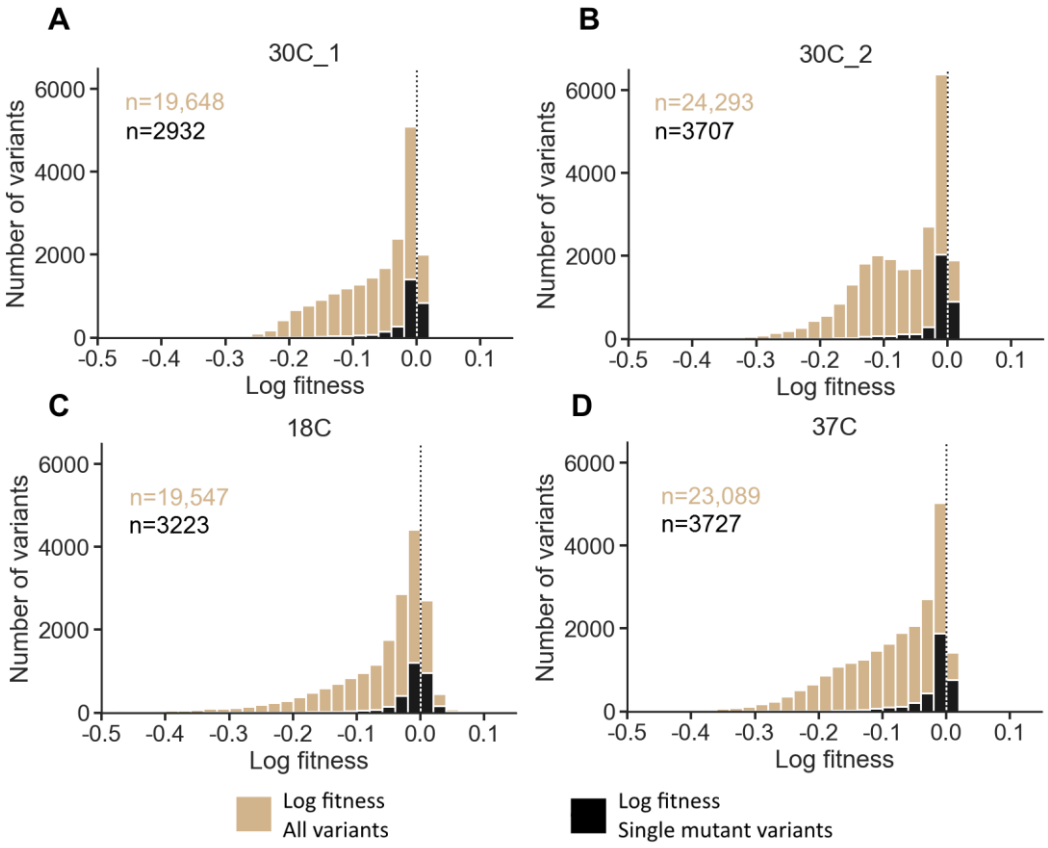


Figure 40, Distribution of fitness effects in different temperature environments

The number of genotypes detected in an experiment (tan) A-D, as well as just single mutant genotypes (black), plotted in 0.02 log fitness bins. The number of values that constitute each group is indicated in its respective colour.

The distribution of log fitness effects at 18°C and 37°C shows a very similar pattern to 30°C (**Figure 40C-D**). The range is comparable, with the tail of the peak ending ~-0.3 as well as the majority of genotypes being at near 0 log fitness. The percentage of deleterious genotypes differs however. While 18C has slightly less at 61%, 37C has a great proportion at 71%. Mutations in structurally important regions may be more tolerated at 18°C, due to increased base pair stability. This could result in previously detrimental genotypes being tolerated. The change in proportion is very minor however. At 37°C the shift is larger, indicating there may be an overall negative log fitness shift. This could be caused by stability being more of a selective factor, leading to many previously neutral genotypes to now compromising U3 snoRNA.

To better resolve the effect of gene-environment interactions I calculated f_i and p_i for each of the experiments. Aggregation of the genotypes to calculate f_i and p_i values will help expose positional differences between the temperature environments. Any f_i or p_i values from 18C or 37C that deviate from the physiological temperature represent positions effected by gene-environment interactions. In **Figure 41** for both f_i and p_i values I have correlated the two 30°C repeats, and then correlated 30C_2 with 18C and 37C. The same genotypes were used to calculate f_i and p_i when correlating two experiments. This was to avoid any genotypes with large fitness effects detected in only one of the experiments creating false positives.

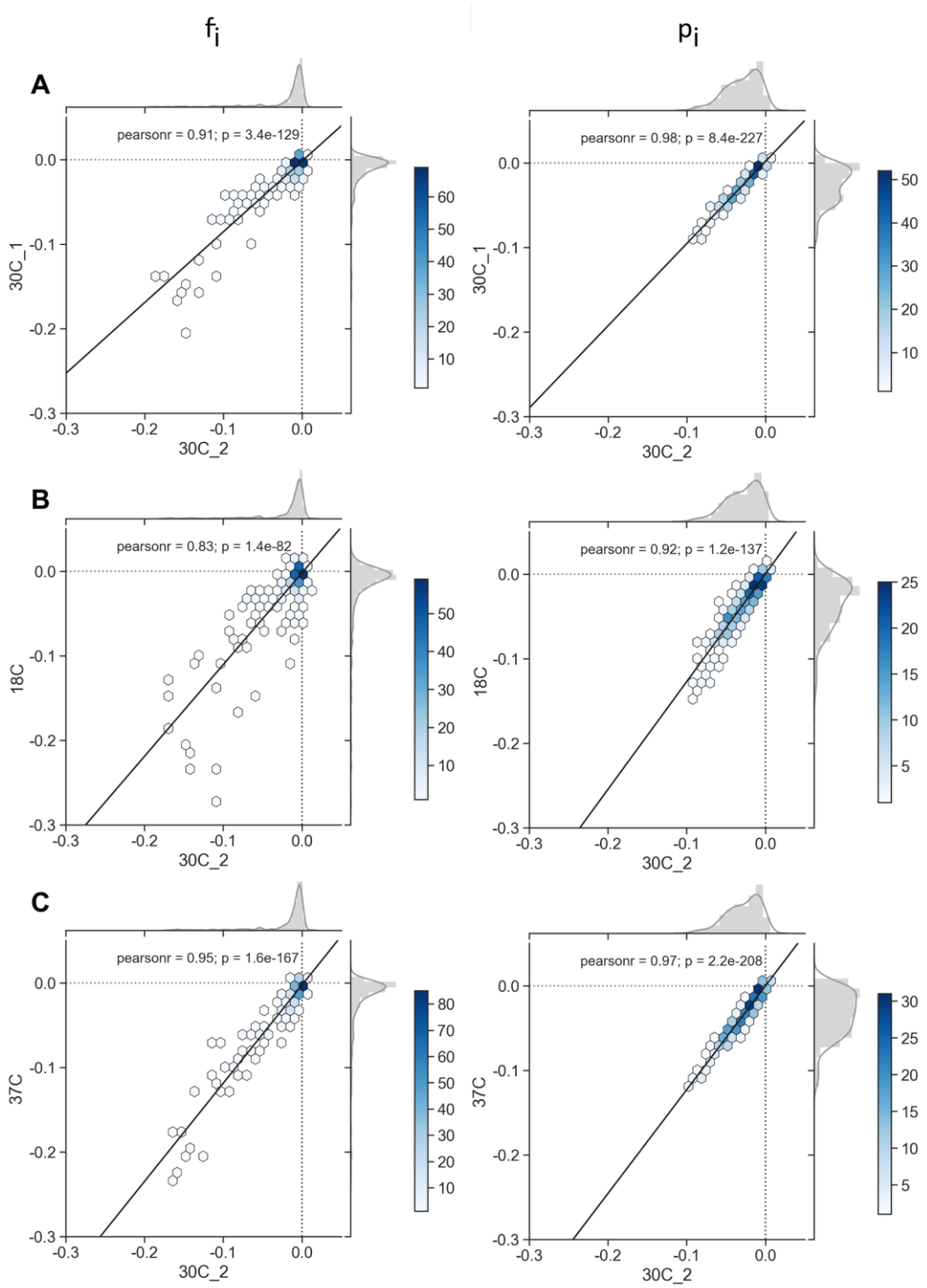


Figure 41, Correlating the f_i and p_i values in different temperature environments

Hexbin density plots between f_i (left column) and p_i (right column) values for different experiments (**A-C**), with regression line and statistics, and distribution plots on the axes. A scale bar for each plot indicates the number of values that constitute each hexagon, with white indicating one value and a darker shade of blue indicating a higher number.

Comparing the two 30°C repeats reveals that when calculating f_i values there is inherent noise when dealing with very low fitness positions located in the box motifs (**Figure 41A-B**). This noise is eliminated when calculating p_i values, by virtue of the ~24,000 genotypes that constitute the p_i values compared to ~3,600 for f_i. The exact number of genotypes is indicated in **Figure 40**.

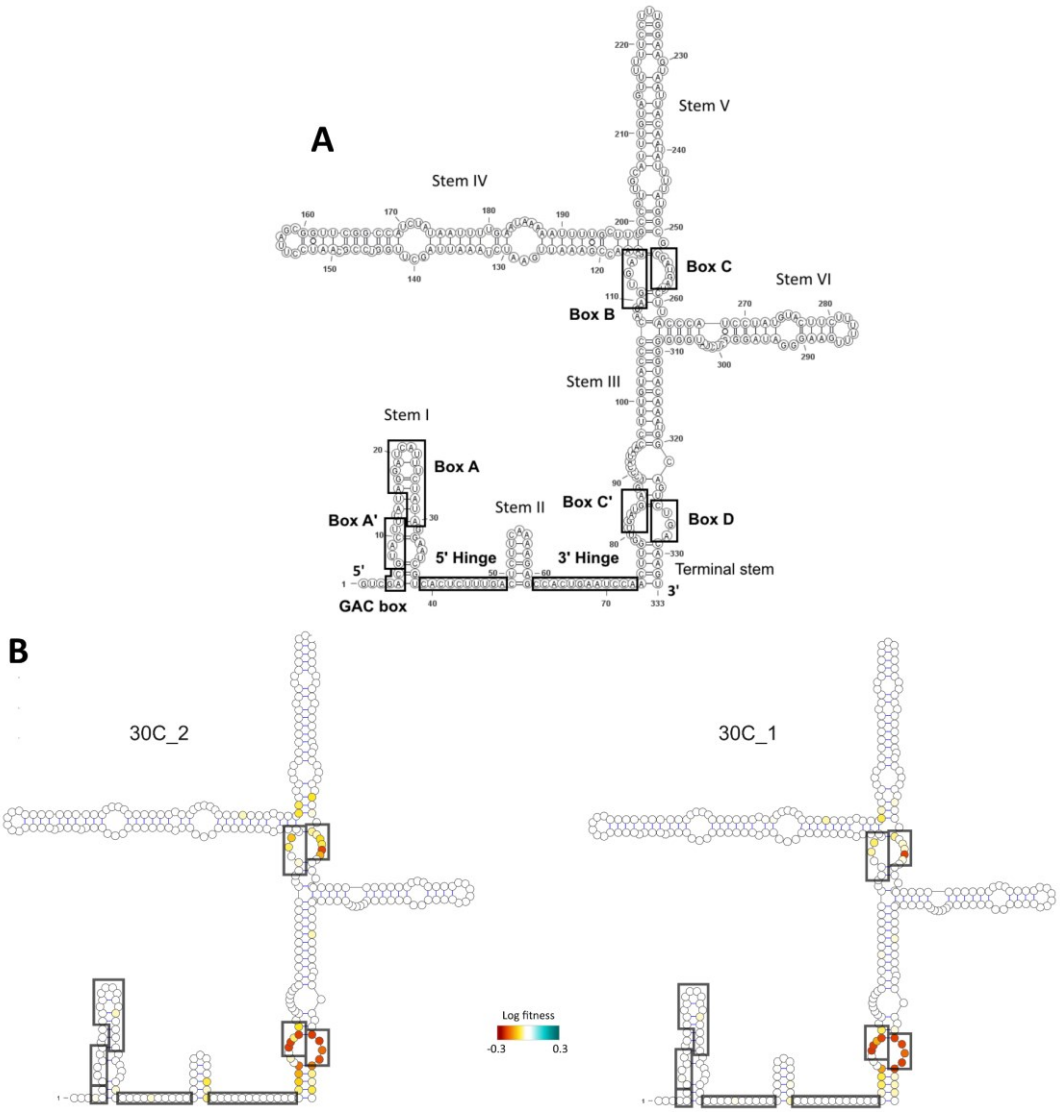
In all temperature environments the majority of f_i and p_i values are at near 0 log fitness, representing U3 snoRNA's robust positions. The remaining positions have a strong linear correlation (**Figure 41A-C**). The f_i values for 18C deviate from 30C_2 the most (Pearson's r = 0.83), with multiple positions having a lower f_i value (**Figure 41B**). This pattern is maintained somewhat with p_i values, with 18C having the weakness p_i correlation (Pearson's r = 0.92). It appears that there are positions with already low fitness values that are exacerbated at 18°C, with intermediate log fitness value positions being less effected. 37C has a stronger correlation with 30C_2 (Pearson's r = 0.95 / 0.97), indicating there are less single positions that deviate considerably. This does not mean there are no gene-environmental interactions at 37°C however, as both 18C and 37C have an overall negative fitness shift. This highlights that at both temperature environments many positions already sensitive to mutation become more so. Any alternations to previously robust positions appear to be subtle.

5.3 Investigating positional gene-environment interactions in a wild type genetic background

To further elucidate the gene-environment interactions identified in **Figure 41**, I have plotted the f_i values for each experiment on the secondary structure of U3 snoRNA in **Figure 42**. For each I have an accompanying plot of f_i values from 30C_2, again calculated using the same genotypes for a fair comparison to the physiological temperature. The colour scale used here is different to that used in (**Figure 27**) to better resolve the smaller effect of gene-environment interactions.

At 30°C the main regions negatively affected by single mutations are the box motifs and base pairs (mainly G-C pairs) in stem II, stem III and stem V as described in **Chapter 4**. There is little variation between the two repeats (**Figure 42B**).

At 37°C base pairs (mainly G-C pairs) in stem structures of U3 snoRNA become more sensitive to mutation (**Figure 42D**). There is a temperature driven reduction in fitness at positions in stem I, the base of stem II, stem III, the base and middle of stem V, base of stem VI and in the terminal stem flanking the box C'/D motif. These are the same positions that became more sensitive when downregulating the proteins of interest, especially Nop58, further reinforcing the structural importance of these positons. At 37°C disruption of these stem structures may have a larger effect on the confirmation of U3 snoRNA, having a larger negative impact on U3's function. Although as noted previously, the relative intensity of gene-environment interactions in this temperature environment are less than the gene-gene interactions described in **Chapter 4**. There is however still an overall reduction to the fitness of G-C


A
Stem V
Stem IV
Box C
Stem VI
Box B
Stem III
Stem I
Box A
Box A'
Stem II
Box C'
Box D
5' Hinge
3' Hinge
Terminal stem
5'
3'
GAC box
B
30C_2
30C_1
Log fitness
-0.3
0.3

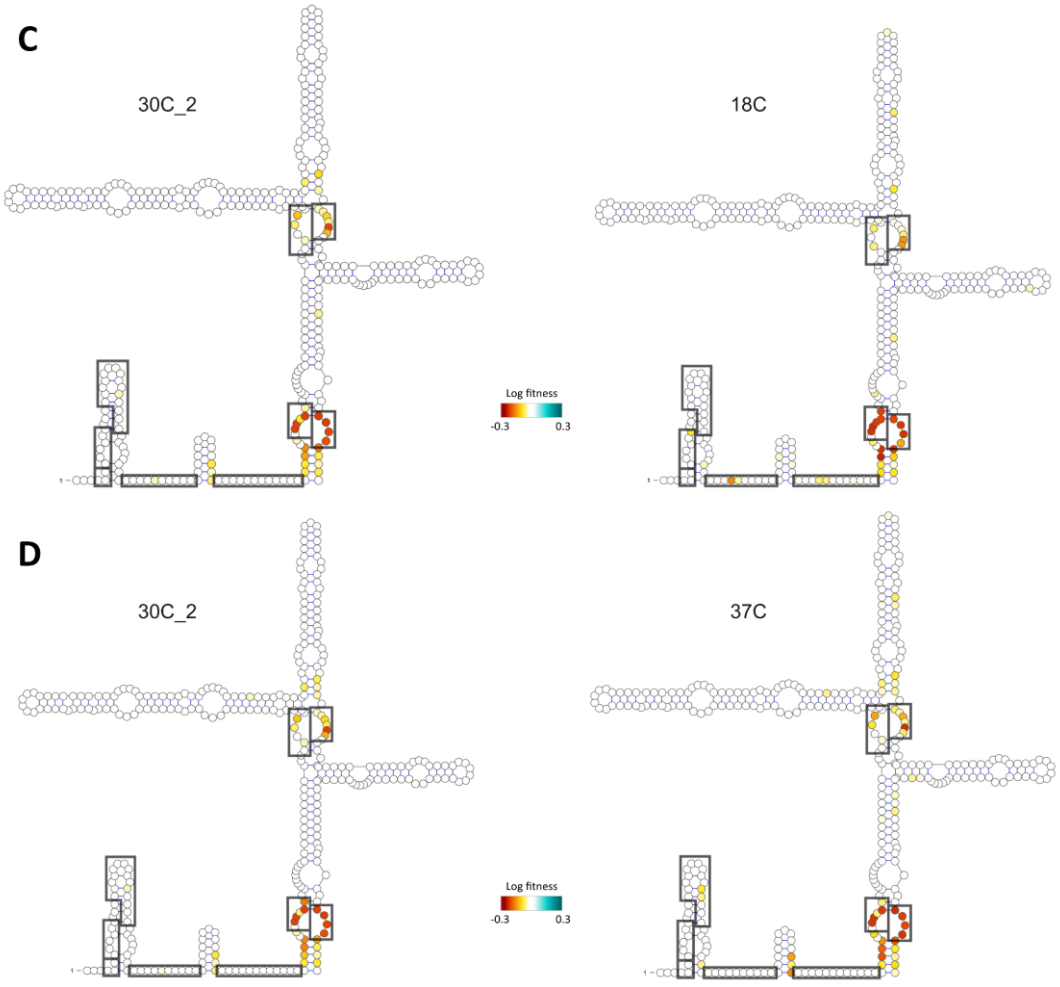


Figure 42, f_i values plotted on the secondary structure of U3 snoRNA in different temperature environments

A. U3 snoRNA secondary structure with nucleotides labelled, box and hinge regions annotated with black boxes, and nucleotide numbers indicated at 50 nucleotide intervals.
B-F. U3 snoRNA secondary structure with f_i values plotted on a colour scale (indicated by colour bar), from negative fitness effects (red) to positive fitness effects (blue), with white indicating a neutral effects. The fi values from the D343_auxin experiment are plotted as a control for each, calculated with the same genotypes.

base pairs (**Figure 43**). At 18°C these same positions show no negative gene-environment interactions apart from base pairs in the terminal stem flanking the C'/D box (**Figure 42C**). In fact base pairs in stem I, stem II and the base and middle of stem V become less sensitive to mutation. This same reduction is seen in 30C_1 however (**Figure 42B**), indicating it could just be a result of variation.

The most affected regions by the change in temperature are the box motifs (**Figure 43**). The box B, box C, box C' and box D motif are all sensitive to mutation at 37°C, while only the box C' is affected at 18°C. Surprisingly the box B/C motif actually becomes more tolerant to mutation at 18°C, coinciding with the increase in tolerance seen at the base of stem V which flanks the B/C motif. Again this same reduction is seen with 30C_1 however, indicating this change may be caused by variation. This region is known to undergo conformational changes after Snu13 binds to facilitate Rrp9 binding (Granneman, et al., 2002). It could be expected that mutations in this region may be better tolerated as they may destabilise the region, compensating for the increased stability and difficulty in making the conformational changes.

It is surprising that at 18°C the box D motif that constitutes the k-turn structure along with C' is not affected (**Figure 43**). Instead at 18°C the box D motif has a very similar log fitness average to growth at 30°C. A more uniform decrease in fitness to the box motifs seen at 37°C. It is not only positions in the box C' motif that are effected at 18°C, but positions flanking it (**Figure 42C**). There are also some select positions at 18°C that appear more deleterious (position 12, 42, 224, 287), however upon further scrutiny these are caused inherent noise in the 18C experiment.

The majority of U3 snoRNA positions are unaffected and remain at wild type levels of fitness when mutated. It appears gene-environment interactions at these temperatures do not compromise the overall robustness of U3 snoRNA.

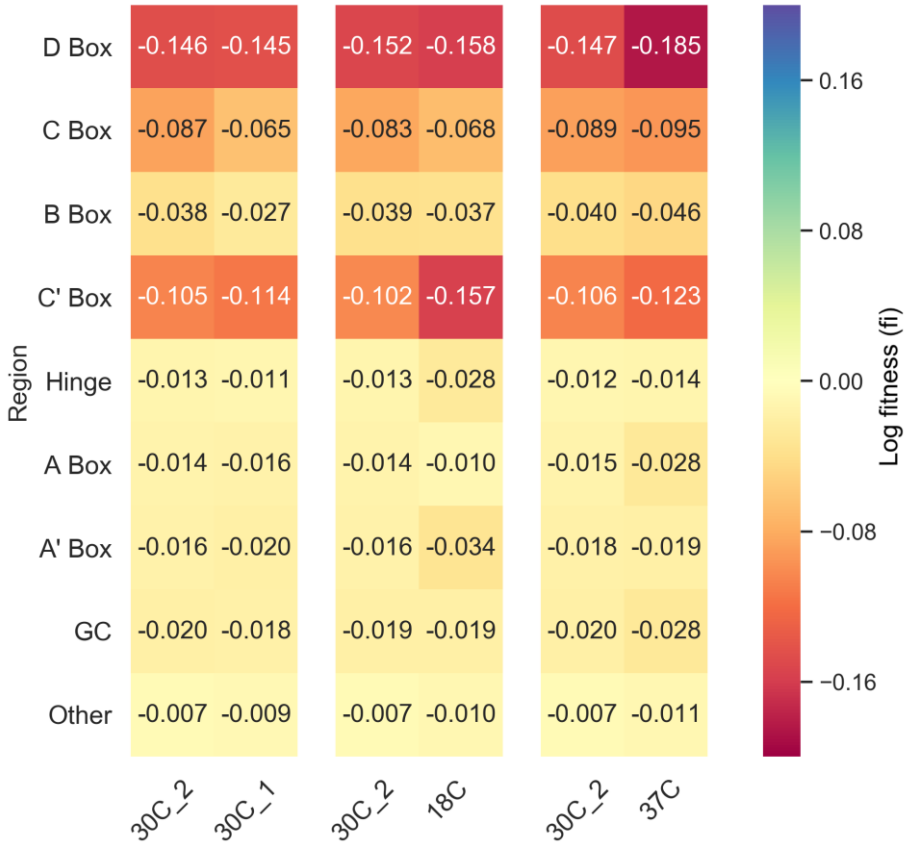


Figure 43, Heatmap for f_i values in different U3 snoRNA regions in different temperature environments

Each row is a U3 snoRNA region (box motifs, hinge regions, GC base pairs and remaining positions), and each column is values from the 30C_2 and indicated temperature environment calculated using the same genotypes. The value is the mean of f_i values in that region, with the colour scale ranging from red for negative log fitness effects, blue for positive log fitness effects and yellow for neutral.

5.4 Effect of gene-environment interactions in all genetic backgrounds

To further explore gene-environment interactions I will now look at all the genotypes that constitute the fitness landscape in each temperature environment. Correlating the f_i values with the p_i values from each temperature environment will expose positional differences caused by a non-wild type genetic background. **Figure 44** shows that in all experiments there are positions that are assigned a negative fitness value when calculating p_i that was neutral when calculating f_i. This is a sign of negative intragenic epistasis and the influence of the genetic background on shaping the fitness effect. The values that are negative in both f_i and p_i have a linear correlation, though not completely linear as the p_i value is normalised to the genetic background.

In 18C and 37C the relationship is mostly altered at positions already negatively affected when calculating f_i (**Figure 44 C-D**). There are however neutral f_i positions that are more deleterious, indicating the presence of intragenic epistasis-environment interactions. The overall negative shift manifests as both deleterious and neutral p_i positions becoming even more negative in a diverse genetic background. The genotypes with mutations in box motifs that were more negative when calculating f_i at 18°C and 37°C are again more negative when calculating p_i, maintaining the linear correlation.

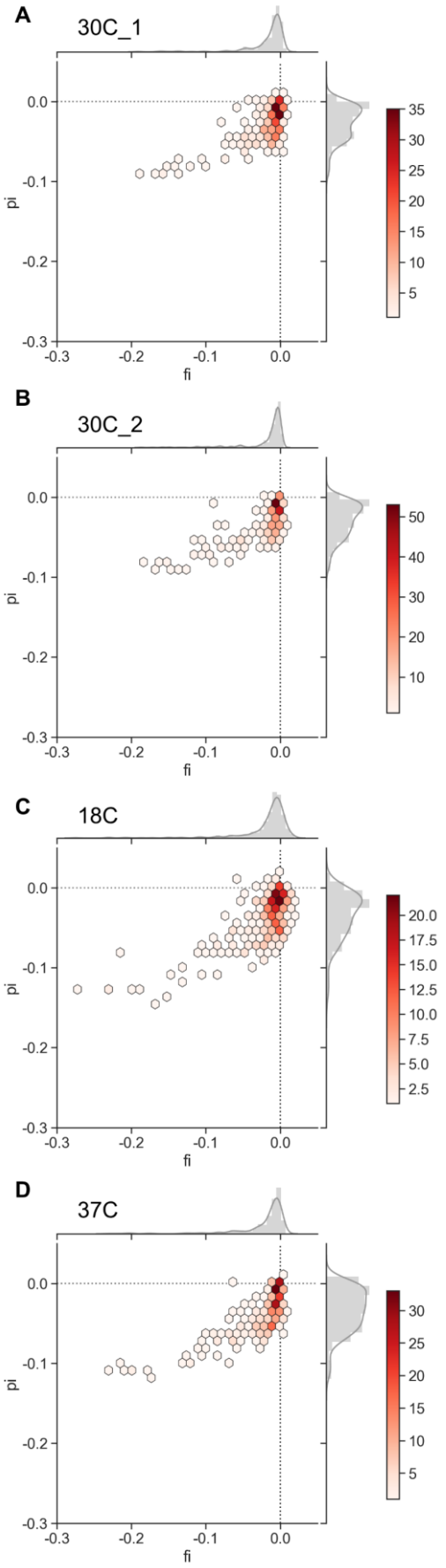


Figure 44, Correlating f_i and p_i values in different temperature environments

Density hexbin plots of f_i values plotted against p_i values for each experiment, with density plots on the axes. A scale bar for each plot indicates the number of values that constitute each hexagon, with white indicating one value and a darker shade of red indicating a higher value.

To look at the fitness implications of mutation number I have plotted the cumulative distribution of fitness effects, grouped by mutation number (**Figure 45A-D**). In all temperature environments as mutation number increases the proportion of genotypes with a negative log fitness value increases. At 18°C the majority of genotypes follow a pattern whereby only low fitness values are extenuated (**Figure 45C**). At 37°C there is a more overall shift seen, with a higher proportion of genotypes being assigned an intermediate lower fitness value (**Figure 45D**). This effect is amplified as mutation number increases. This is indicative of a higher level of destabilisation at 37°C, with high mutation number genotypes being generally more deleterious. At 18°C the gene-environment interactions appear more clustered at specific sites.

The pattern of fitness decreasing as a factor of mutation number is reflected in all temperature environments. To see if this is the cumulative effect of the single mutants that constitute these genotypes or additionally the effect of intragenic epistasis, I have plotted the measured fitness effects against a model with no epistasis (**Figure 45E-H**). This no epistasis model is calculated by aggregating the log fitness effects of single mutants for each of the different multiple mutant genotypes. In all temperature environments the measured effect is lower than the no epistasis model, reinforcing the presence of intragenic epistasis. As noted in **Chapter 4** post six mutations the no epistasis model becomes lower than measured effects due to a saturation effect, whereby after a certain threshold U3's function cannot be further reduced. At 37°C (**Figure 45H**) the same pattern can be seen as with the cumulative distributions (**Figure 45D**), with log fitness strongly decreasing as a factor of mutation number. At 18°C there is more variation than at 30°C, but the median log fitness is very similar to 30°C as mutation number increases.

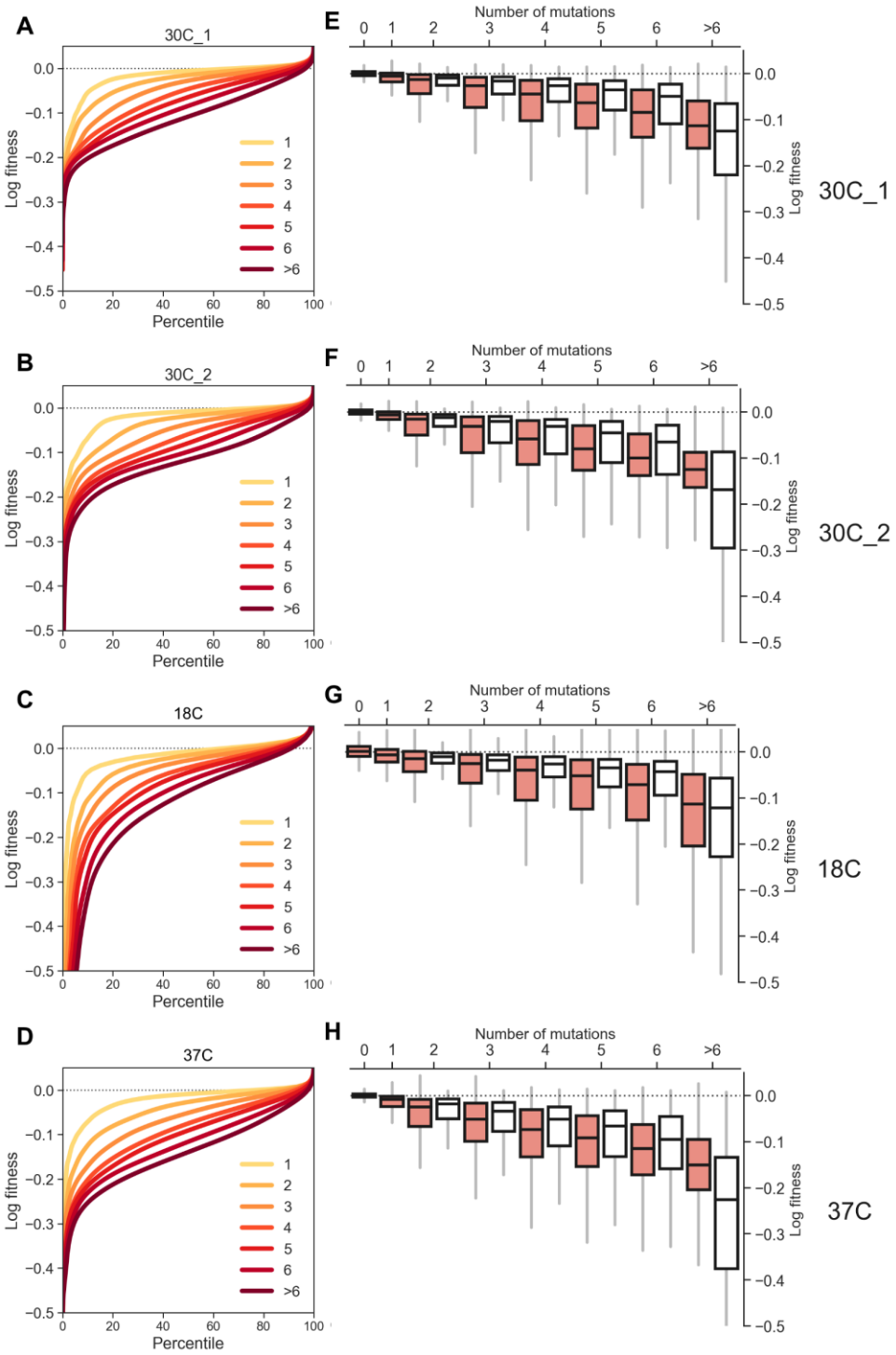


Figure 45, Effect of multiple mutations in different temperature environments

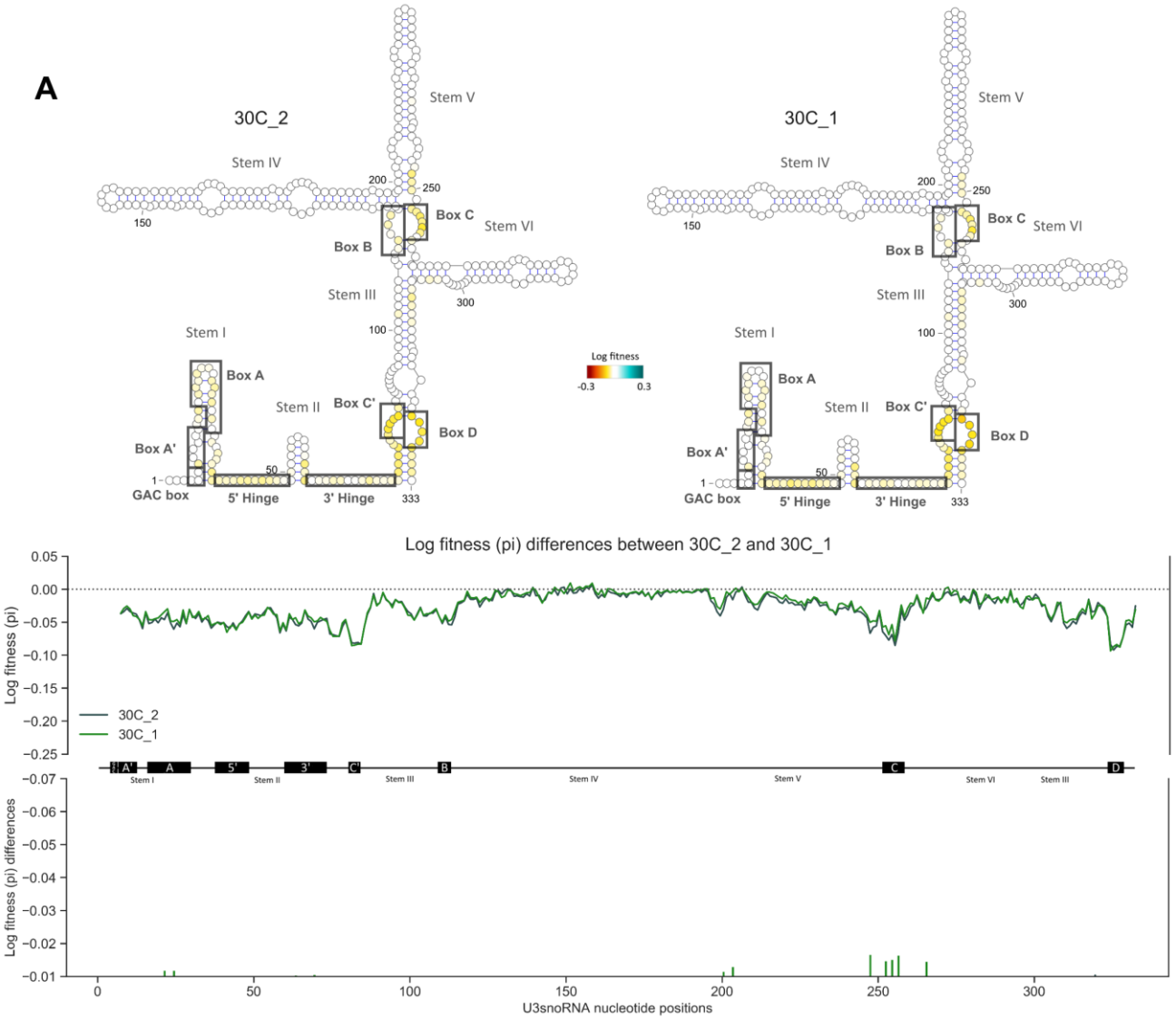
A-D. Plots for each experiment, showing the cumulative distribution of fitness effects for genotypes, grouped by the number of mutations present.
E-H. Box plots showing for different numbers of mutations the mean measured log fitness effects (orange) compared to aggregation of the equivalent single mutant genotypes to represent a model with no epistasis (white). The boxes show the median, inter-quartile range, and upper and lower extremes.

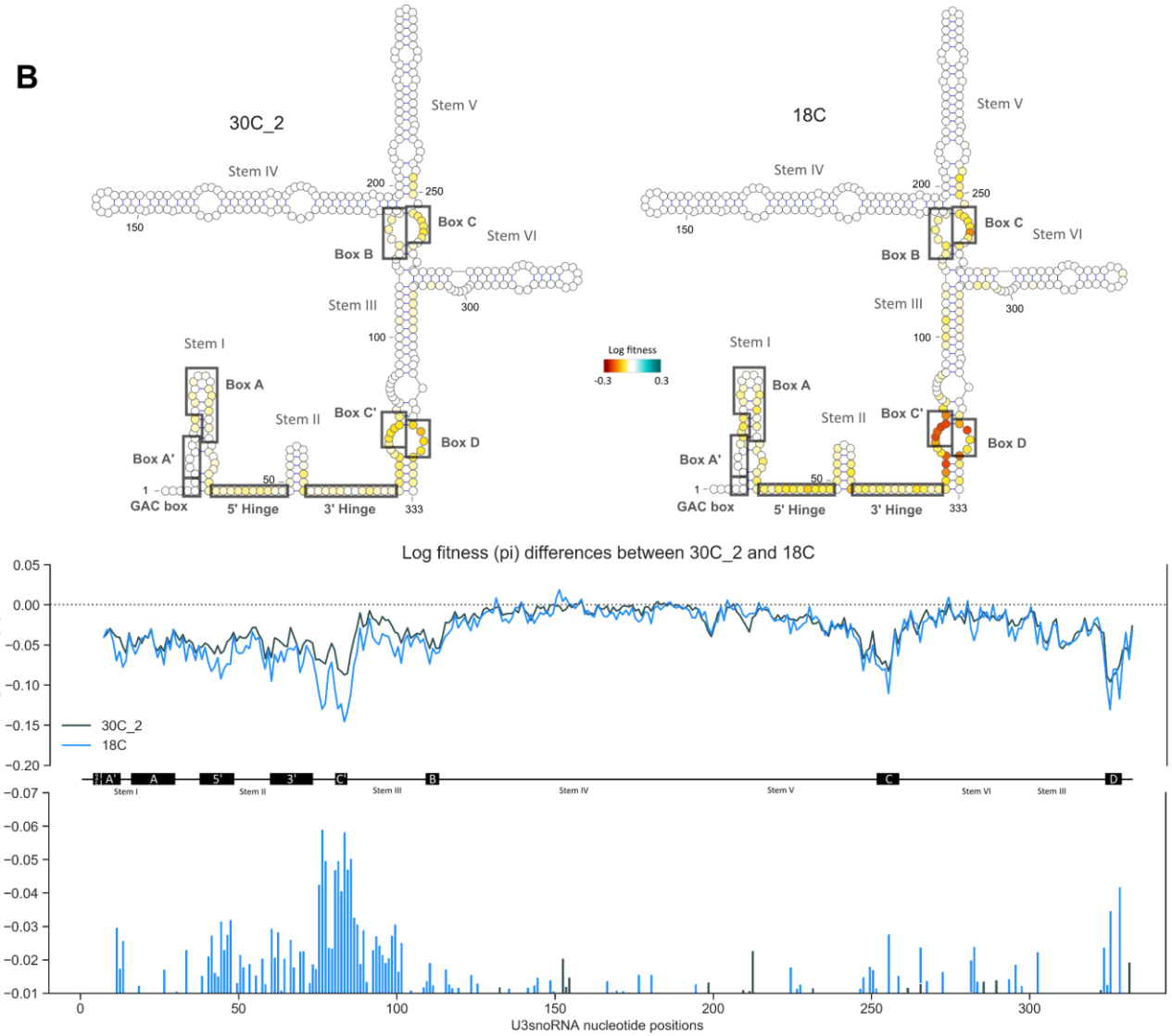
5.4.1 Positional effect of gene-environment interactions

To investigate exactly which positions are affected when incorporating multiple mutations to calculate p_i, and ascribe a more molecular explanation, I have plotted the p_i values on the secondary structure of U3 snoRNA (**Figure 46**). Again the colour scale used here is different to that used in **Figure 34** to better resolve the smaller effect of gene-environment interactions.

At 30°C in addition to the regions identified when calculating f_i, regions in stem I, stem III and the hinge regions become sensitive. This pattern at the physiological temperature is covered in **Chapter 4**. Both 30°C repeats have very similar fitness landscapes (**Figure 46A**).

At 18°C the most striking gene-environmental interactions appear to be at the box C' motif (**Figure 46B**), as seen when looking at f_i values (**Figure 42C**). Mutations in the box C' motif and the flanking base pairs are significantly more deleterious at 18C, compared to 30C. The box C, box B and box D motifs are also now effected when calculating p_i (**Figure 47**), which wasn't seen with f_i (**Figure 43**), the effect at the box C' motif is the most significant. This could either be due to difficulty in making the conformational changes that occur as proteins bind this region, or obscuring a protein's ability to recognise this region. There also appear to be gene-environment interactions at regions that form duplexes with the pre-rRNA. The A/A' and hinge regions are more sensitive to mutation at 18°C. Stem III and the base of stem VI is also effected, which may be disrupting an interaction between nucleotides 304-315 of U3 snoRNA the 18S pre-rRNA (Kudla, et al., 2011). With colder temperatures perhaps more correct complementarity is required to avoid aberrant duplex formation.


A
30C_2
30C_1
Stem V
Stem IV
Stem VI
Box C
Box B
Stem III
Stem I
Box A
Box A'
Stem II
Box C'
Box D
GAC box
5' Hinge
3' Hinge
Log fitness
-0.3
0.3
Log fitness (pi) differences between 30C_2 and 30C_1
Log fitness (pi)
Log fitness (pi) differences
U3snoRNA nucleotide positions


B
30C_2
18C
Stem V
Stem IV
Stem VI
Stem III
Stem I
Stem II
Box A
Box A'
Box B
Box C
Box C'
Box D
GAC box
5' Hinge
3' Hinge
Log fitness
-0.3
0.3
Log fitness (pi) differences between 30C_2 and 18C
Log fitness (pi)
Log fitness (pi) differences
U3snoRNA nucleotide positions
30C_2
18C

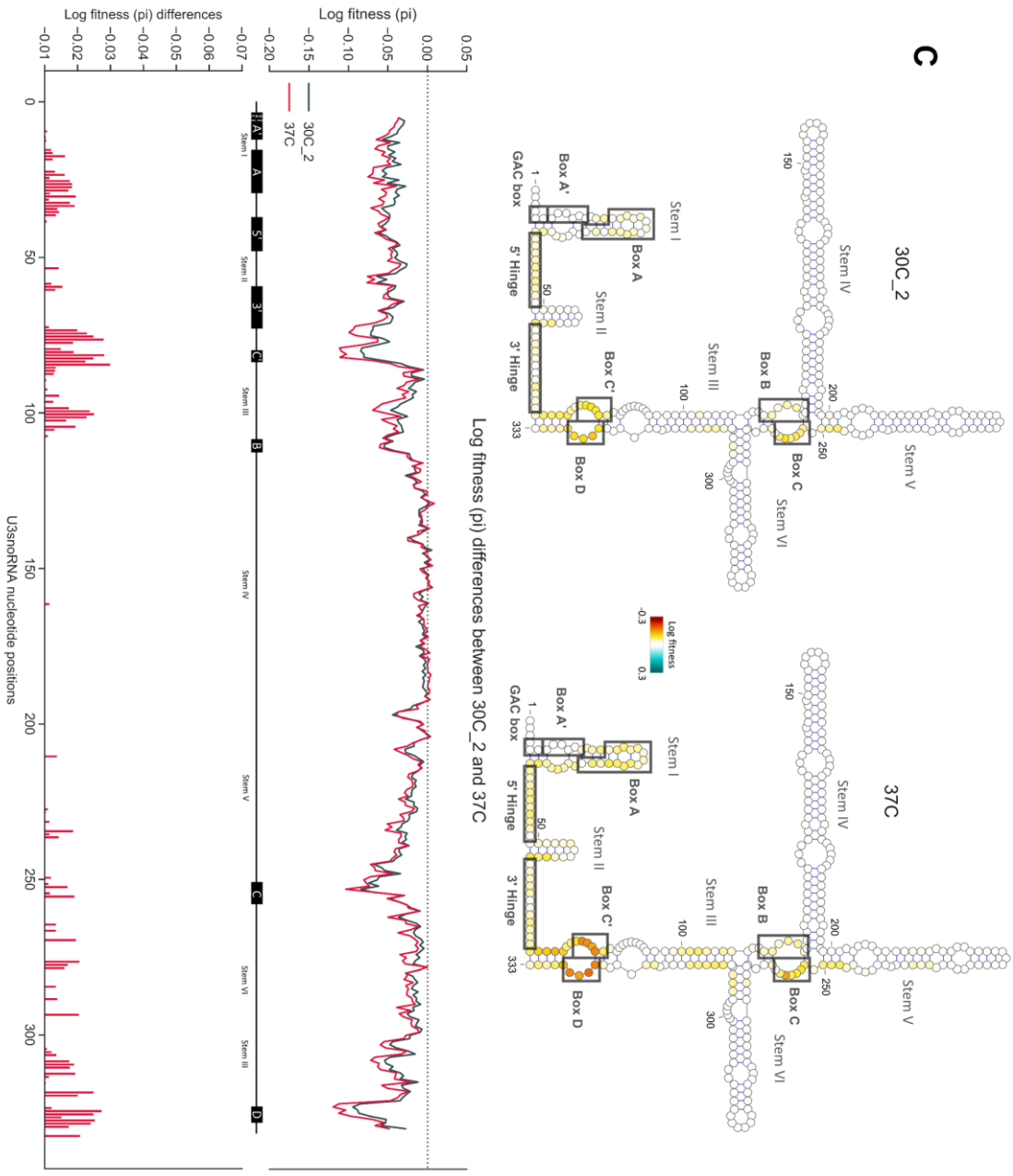


Figure 46, Comparing p_i values from different temperature environments

U3 snoRNA secondary structure with p_i values plotted on a colour scale (indicated by colour bar), from negative fitness effects (red) to positive fitness effects (blue), with white indicating neutral effects. The p_i values from D343_auxin and the protein of interest strain are calculated with the same genotypes, with the same p_i values from both being plotted on the secondary structure of U3 snoRNA, in a line chart, and the difference between the p_i values for each position is plotted in bar chart. The characterised regions of U3 snoRNA (A box, A' box, hinge regions, C box, B box, C' box, D box) are indicated in a schematic between the line and bar chart. U3 snoRNA nucleotide positions are indicated in 50 nucleotide intervals on the bar chart and on the U3 snoRNA secondary structure.

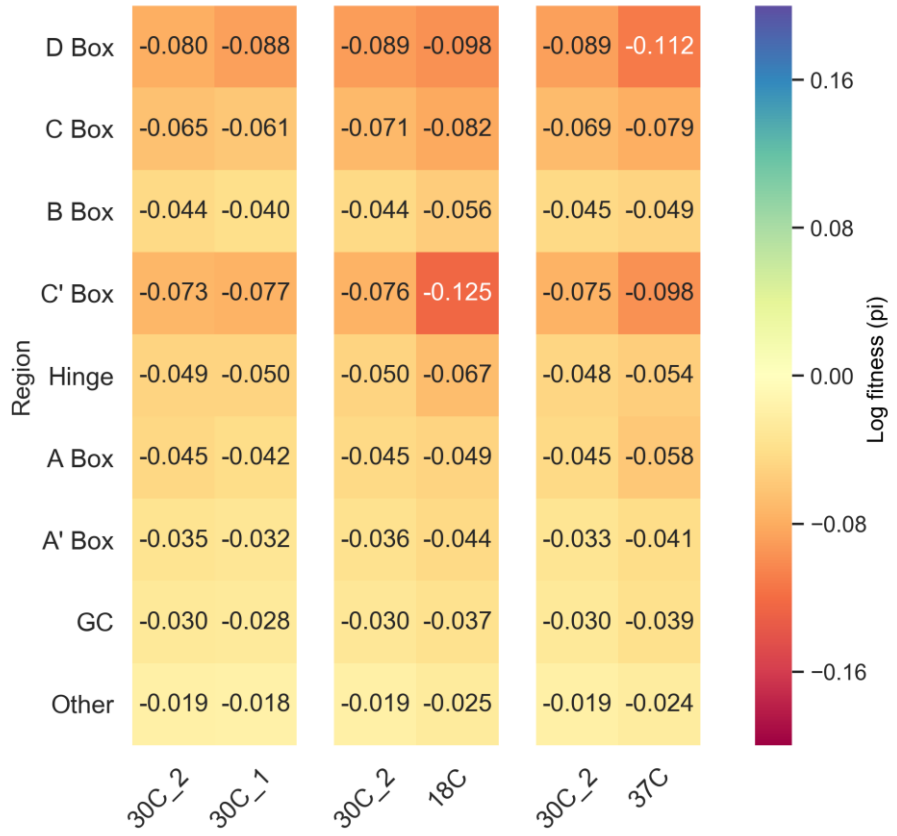


Figure 47, Heatmap for p_i values in different U3 snoRNA regions in different temperature environments

Each row is a U3 snoRNA region (box motifs, hinge regions, GC base pairs and remaining positions), and each column is values from the 30C_2 and indicated temperature environment calculated using the same genotypes. The value is the mean of p_i values in that region, with the colour scale ranging from red for negative log fitness effects, blue for positive log fitness effects and yellow for neutral.

I originally thought that mutations in the pre-rRNA binding regions could be beneficial as they compensate for the increased difficulty in the formation and disassembly of the pre-rRNA-U3 snoRNA duplex, but this does not seem to be the case. Unlike with the f_i values, with p_i it initially appears that locations important to stability are now affected. These include stem I, stem II, stem III and base of stem V. It may not be stability related however as stem I and stem II undergo conformational changes to facilitate binding to the pre-rRNA. These stems are also known to be essential for recognition by the Mpp10 complex (Wormsley, et al., 2001), which is also essential for disrupting these stems, and so with multiple mutations it may impair recognition by the Mpp10 complex. While plasticity is essential in the U3 snoRNA secondary structure, certain genotypes could result in competing secondary structure conformations that are not viable at 18°C.

At 37°C stem I, stem II and stem III are affected to a greater extent than at 18°C, and in addition the base of stem VI and stem V are (**Figure 46C**). These are all GC rich structurally important stems. The extent of the fitness reduction is much more uniform, a similar pattern but to a much lesser extent to that seen when Nop58 is downregulated (**Figure 34F**). The box C and box B motifs are affected similarly to at 18°C (**Figure 47**), with the box D motif being most affected, as seen with f_i values (**Figure 43**). The hinge region shows very little divergence from 30°C, indicating duplex formation is not impacted. These results continue to highlight the increased destabilisation at 37°C. Overall the fitness effects are more subtle, which is expected as there is only a small growth defect when culturing at 37°C due to the heat shock response (**Figure 39**).

When comparing the p_i values of positions located in different base pair types (**Figure 48**), it reinforces the impact of the different temperature environments.

The two 30°C repeats show no significant differences. At 18°C positions in the conserved box motif and hinge regions are most affected (p-value = $3.3x10^{-5}$), while at 37°C it is G-C pairs (p-value = 0.027). At 18°C two outliers G-C and A-U/W pair positions can be seen, these however are flanking the C'/D motif, so are in reality still related to the box motifs. At 18°C G-C pairs are actually the least impacted, perhaps due to an increase in plasticity being less detrimental and even beneficial. Unpaired and A-U/W pairs are effected at 18°C, which was not seen when calculating f_i, as multiple mutations may provide great scope incompatible secondary structures. These results again highlight that at 18°C the strongest gene-environment interactions perturb protein binding, while destabilisation of U3 snoRNA is the strongest interaction at 37°C.

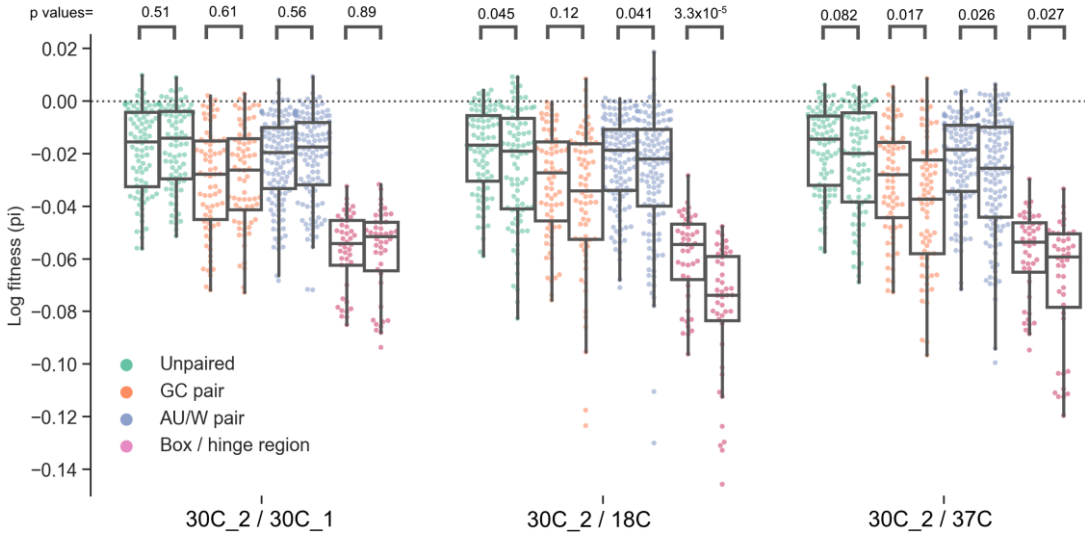


Figure 48, Box plot comparing different position types in different temperature environments

Comparing the p_i values from different experiments in positions that are either unpaired (green), a GC pair (orange), an AU / wobble base pair (blue) or in the box motif / hinge region (red). Student's t-test p-values between experiments are indicated.

5.4.2 Alignment with U3 snoRNA from different environments

While the D343 strain I have used grows optimally at 30°C this is not the case for all strains of Saccharomyces cerevisiae. In industry processes uses such as bioethanol production (Mohd Azhar, et al., 2017) or wine production (Pretorius, 2000) a strain with robust thermotolerance is essential. Due to the low temperatures used in the fermentation of wine, hybrids with the more cryotolerant S. *kudriavzevii* are also used (González, et al., 2006). The genomic contribution of *S. kudriavzevii* that increases cryotolerance in the hybrid include increased expression of Nsr1 (involved in pre-rRNA processing), and Gut2 and Dpd1 (involved in glycerol synthesis) (Ortiz-Tovar G, 2018).

Since expression increases in genes involved rRNA synthesis and ribosomal proteins during the cold shock response (Sahara, et al., 2002), I thought it would be interesting to look at the snR17a gene (U3 snoRNA) in more thermotolerant strains. In **Table 6** I have listed with description a selection of *S. cerevisiae* strains, along with a *S. kudriavzevii* strain. In **Figure 49A** I have aligned the snR17a gene from the D343 strain with said strains. First looking at the 13 *S. cerevisiae* strains, the majority of the SNR17a sequence is unchanged. If any major changes to U3 snoRNA exist, they may be in snR17b.The two most common substitutions that are present in snR17a are G->A at position 214, removing an internal loop from stem V and creating a A-U base pairing, and a C->G in a 5 nucleotide bulge in stem VI. The substitution at position 214 is present in 6 strains and creates an additional base pairing in an area important to both stability and plasticity. In **Figure 49B** I have plotted the log fitness effects for all substitutions at position 214 for all temperature environments for both single, and single and low fitness double mutants. The substitution to A appears to have no fitness impact. There appears to be a slight negative fitness impact with a C substitution at 37°C, but it is very small.

Table 6, Strains with thermotolerance

Strain	Strain description (from https://www.yeastgenome.org/)	Reference
AWRI796	A South African wine strain that ferments more successfully at warmer temperatures and is more suited to the production of reds.	Borneman AR, et al. (2011)
AWRI1631	Australian wine yeast, a robust fermenter and haploid derivative of South African commercial wine strain N96.	Borneman AR, et al. (2008) PMID:18778279
CBS7960	Isolated from a cane sugar bioethanol factory in Sao Paulo, Brazil.	
EC1118	Probably the most widely used wine-making strain worldwide based on volume produced. There are >100 genes present in S288C that are missing from EC1118.	Novo M, et al. (2009)
JAY291	Non-flocculent haploid derivative of Brazilian industrial bioethanol strain PE-2; it produces high levels of ethanol and cell mass, and is tolerant to heat and oxidative stress. JAY291 contains well-characterized alleles at several genes of known relation to thermotolerance and fermentation performance.	Argueso JL, et al. (2009)
Kyokai7	Kyokai No. 7 (K7) is the most extensively used sake yeast, and was first isolated from a sake brewery in Nagano Prefecture, Japan, in 1946.	Akao T, et al. (2011)
LalvinQA23	QA23 is a cold-tolerant Portuguese wine strain from the Vinho Verde region.	Borneman AR, et al. (2011)
PW5	PW5 was isolated from fermented sap of a Raphia palm tree in Aba, Abia state, Nigeria in 2002.	Nijkamp JF, et al. (2012)
T7	T7 was isolated from oak tree exudate in Missouri's Babler State Park.	
VIN13	VIN13 is a cold-tolerant South African wine strain.	Borneman AR, et al. (2011)
VL3	VL3 was isolated in Bordeaux, France, and is most suited to the production of premium aromatic white wines with high thiol content (citrus and tropical fruit characters).	Borneman AR, et al. (2011)
YJM789	YJM789 is the haploid form of an opportunistic pathogen derived from a yeast isolated from the lung of an immunocompromised patient in 1989. YJM789 is useful for infection studies and quantitative genetics owing to its divergent phenotype, which includes flocculence, heat tolerance, and deadly virulence.	Tawfik OW, et al. (1989)
ZTW1	ZTW1 was isolated from corn mash used for industrial bioethanol production in China in 2007.	
CR85	A *Saccharomyces kudriavzevii* strain used in production of alcoholic beverages, including lager beers and pinot noir wine. It is isolated widely from the bark of Oak trees. *S. kudriavzevii* is adapted to grow at much colder temperatures	

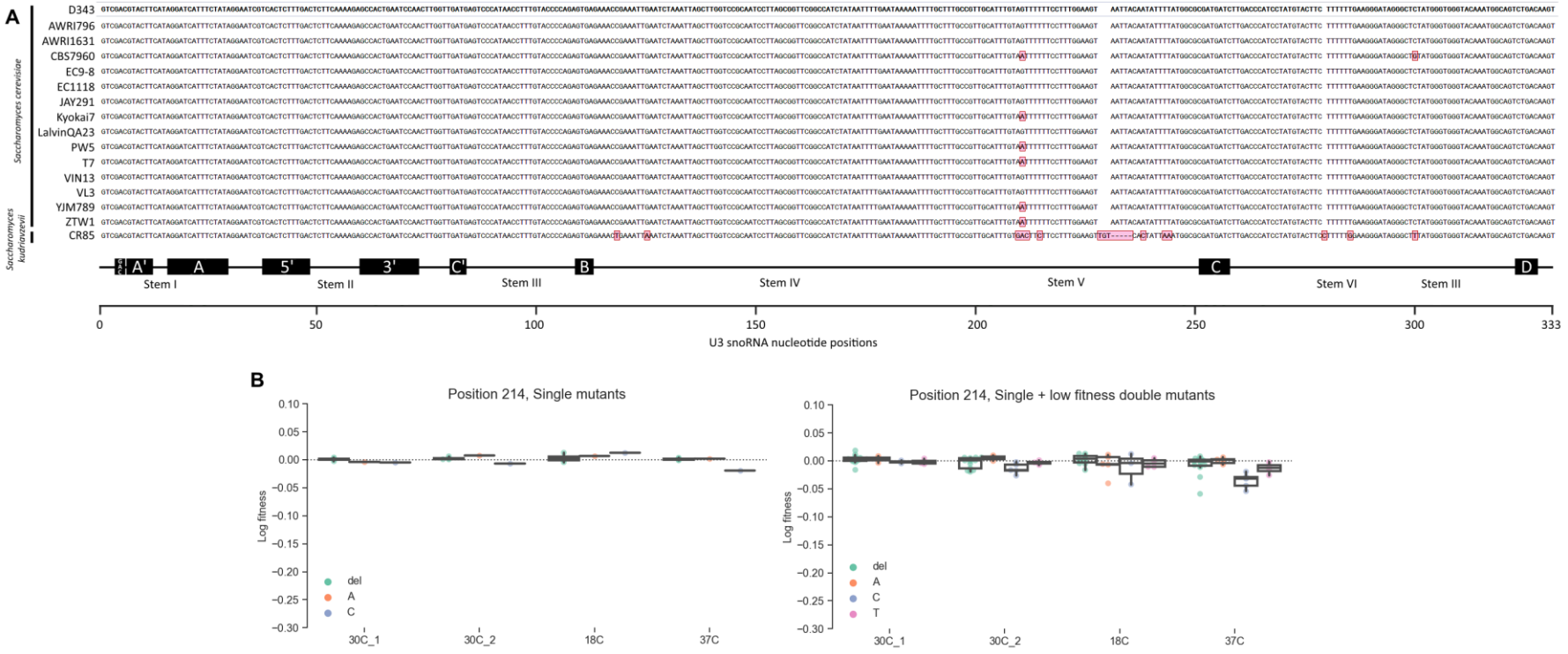


Figure 49, Alignment of SNR17a from different yeast strains

A. Alignment of 14 Saccharomyces strains, with the strain name and species indicated. Differences to the D343 strain are annotated in red. Below the sequences the conserved regions of U3 snoRNA are indicated as well as the nucleotide positions. **B.** Box plot showing the difference in fitness effect caused by different mutations (substitutions from G to T, A or C and single deletion) at position 214 in 30C_1, 30C_2, 18C and 37C experiments. The values are from single mutant genotypes or single and double mutant genotypes with a low fitness impact background.

The snR17a gene in CR85 (*S. kudriavzevii*) has more extensive differences. This consists of:

- C->T substitution at position 120 disrupting a GC base pair between positions 120-194 in stem IV.
- G->A substitution at position 127 in a 3 nucleotide bulge in stem IV.
- Series of substitutions and deletions in stem V resulting in additional base pairing and different internal loops.
- C inserted at position 281 and a T->G substitution at position 286 creating an additional GC base pair flanking the hairpin loop of stem VI.
- C->T substitution at position 301 in a 5 nucleotide bulge of stem VI.

The extensive changes in stem V and base of IV may have to do with the conformational changes that occur in this area when Snu13 binds the box B/C motif. These changes may compensate for the hyperstability caused by low temperatures. When looking at the f_i values at 18°C this region was more tolerant to mutation, however this same pattern was seen in 30C_1, so it is uncertain if this truly represents suppressor mutations. The same G->A substitution present in the six *S. cerevisiae* strains at position 214 is also part of the stem V changes in *S. kudriavzevii*. A genotype matching the U3 snoRNA found in *S. kudriavzevii* does not exist in the library, but it would interesting to see the fitness effect if this genotype was created.

In stem VI the additional GC pair seems to be creating additional stability in a region functionally unimportant to U3 snoRNA, as it readily tolerates mutation. The C->T substation at position 301 is in an equally robust position, but is at the same position as a C->T substitution in strain CBS7960, indicating it may have a role.

5.5 Discussion

A fitness landscape maps the fitness implications of different genotypes in a given environment. Gene-environment interactions are known to have a large influence on the consequence of mutations. A beneficial allele in one environment may be detrimental in another. My deriving the fitness landscape of a gene in different environments, gene-environment interactions can be systematically characterised.

Upon comparing the fitness landscape of U3 snoRNA in three different temperature environments, I have found many temperature specific differences. Gene-environment interactions are enriched in functionally important regions. These appear to be a result of perturbations to the thermodynamic stability and secondary structure of U3 snoRNA, by either destabilisation at 37°C or a hyperstability at 18°C. These interactions are most prevalent when incorporating multiple mutations into the analysis. The robustness of U3 snoRNA is broadly maintained in different temperature environments when faced with single mutations. The addition of multiple mutations into the genetic background of a focal mutation however leads to more deleterious fitness effects. This is similar to the pattern seen with gene-gene interactions, that multiple mutations render U3 snoRNA more susceptible to interactions. It appears that there is a more scope for gene-gene interactions to influence the U3 snoRNA fitness landscape, as the gene-environment interactions are much less severe, and temperatures beyond those I tested are not viable. The D343 strain is no longer active past 39°C. This may be due to molecules having evolved to withstand temperature changes. The gene-gene interactions I tested are also much more specific than the ubiquitous change caused by an altered temperature environment.

5.5.1 Gene-environment interactions in different temperature environments

By completely the assay at higher temperatures destabilising mutations had increased potency, exposing nucleotides in stem structures vital to maintaining structural integrity. Conversely at low temperatures, conserved protein binding regions become more sensitive to mutation. It may be that mutations in regions requiring conformational flexibility are also affected.

At 18°C the strongest gene-environment interactions impact mutations within and surrounding the box C' motif. This effect is maintained whether looking at single or multiple mutations. A possible explanation could be due to the conformational changes that occur at this region. Lack of required base pairs and addition of aberrant base pairing may both obscure this site from being recognised during assembly and be harder to maintain the required conformation. The fact the effects are clustered on the 5' side of the stem, with the juxtaposed box D motif unaffected, indicates it could be due to protein recognition and not the maintenance of the k-turn motif. This effect is not seen at the box B/C motif, indicating conformational changes here may rely more on protein remodelling. Multiple mutations are required for the secondary structure of U3 snoRNA to become less tolerant of perturbation. This may be due to destabilisation, but it is known that cold sensitive mutations in RNA structures can be due to folding into an alternate conformation. Mutations in the structurally important stem structures may restrict essential plasticity. This is highlighted by mutations in G-C base pairs being the least affected in the 18°C temperature environment, which would be unexpected if the molecule was destabalised. Mutations in the hinge region are less tolerated, with the increased free energy at 18°C perphaps increasing the likelihood of forming incorrect duplexes.

In the 37°C temperature environment the preservation of structural important stems and subsequent correct U3 snoRNA secondary structure has increased importance. This is due to the destabilising effect of mutations in these regions being exacerbated. This is seen in a position independent manner by the log fitness of genotypes decreasing as the number of mutations per genotype increases. There are also specific positions effected, located in G-C rich stem structures, exposing their importance to correct RNA folding. 37°C does not appear to be enough of a temperature shift to disrupt duplex formation as the hinge regions are unaffected. Overall the scale of gene-environment interactions are lowest at 37°C. This was expected considering only a small growth defect was observed, in some part due to the heat shock response. Since many gene expression changes due to the heat shock response cluster in ribosome biogenesis pathways, it is to be expected these pathways including U3 snoRNA are well adapted to changing temperature environments.

5.5.2 Topology of the fitness landscape in different temperature environments

It was found that changes in the environment could provide a bridge across low fitness valleys (Steinberg & Ostermeier, 2016). With U3 snoRNA however it appears that there are no positions that become less sensitive to mutation to provide movement in sequence space. There may potentially be positions surrounding the box B/C motif at 18°C, however due to similar changes being seen between 30°C repeats, this is not certain. There may also be specific genotypes in all three temperature environments that display positive gene-environment interactions. Genotypes can be maintained if they are beneficial in one environment, but neutral to detrimental in another, as long as the environment fluctuates enough (Qian, et al., 2012). In order to identify these it would require using the wild-type genotypes to calculate estimate noise as a

function of coverage (as discussed in **4.5.4**), as single variants can be unreliable. The changes at 18°C and 37°C to the topology of the fitness landscape compared to 30°C appear to restrict evolutionary trajectories. The plateau of neutral mutations in the robust regions of U3 snoRNA is reduced. The sides of the peak caused by mutations in conserved regions are steepened as mutating these regions is less tolerated.

Many of the alterations caused by the heat and cold shock responses are involved in ribosome biogenesis, indicating that molecules involved in these processes such as U3 snoRNA are likely more robust to changes in temperature environments. Deriving the fitness landscape of a different molecule may expose higher sensitivity to gene-environment interactions. This may also provide an opportunity to find positive interactions if the molecule is less robust to temperature changes.

5.5.3 Potential future work

To confirm that it is perturbations to the secondary structure or stability of U3 snoRNA that are behind the gene-environment interactions observed, the thermodynamics of deleterious genotypes could be studied. This could complement the observation that an increased mutation number leads to relatively higher proportion of deleterious genotypes at 37°C. This could also help explain gene-environment interactions surrounding the C' box at 18°C.

While temperature is one of the most pervasive changes than an organism can face, there are additional environments that could be tested. DMSO affects RNA function and structure (Lee, et al., 2013). While DMSO had a small influence on the fitness landscape of a trna (Li & Zhang, 2018), the series of

proteins and subsequent conformational changes that occur with the U3 snoRNP may result in a larger effect. This may highlight the same structurally important stems identified at 37°C. A lower pH is also known to affected RNA structure (Bina-Stein & Crothers, 1974), providing a similarly interesting environment to test.

The available nutrients to yeast are strongly linked with growth and thus ribosome biogenesis. A yeast approaching stationary phase has a much lower requirement for protein synthesis. Of the nitrogen sources available to yeast, proline results in the slowest growth (Huang & Brandriss, 2000). This is despite the fact proline is the most abundant nitrogen source in locations such as grapes where yeast may be found (Bisson, 1991). Completing the competition experiment with proline as the nitrogen source will therefore reveal the fitness landscape of U3 snoRNA in an environment where its function is in lower demand. It is also known that a nutrient change can alter the pre-rRNA processing pathway, favouring the non-productive A_3 pathway to halt ribosome synthesis (Kos-Braun & Koš, 2017). Having proline as the nitrogen source may lead to many positive gene-environment interactions, as many genotypes will be tolerated with a lower demand for ribosome biogenesis.

6 Concluding remarks

Using a high-throughput fitness assay I have found gene-gene and gene-environment interactions to be prevalent in the fitness landscape of U3 snoRNA. These appear exclusively detrimental, perturbing function through a multitude of mechanisms, and restricting evolutionary trajectories. Positions sensitive to mutation and intragenic epistasis in physiological conditions are vulnerable to further disruption from gene-gene and gene-environment interactions. Even positions previously displaying genetic robustness can be compromised, indicative of the inherent robustness threshold being lowered. The assay appears a good candidate to characterise these interactions, providing the potential to ascribe molecular mechanisms.

The conditions I have used to probe genetic interactions exposed U3 snoRNA to a broad range of adversities. The positions of U3 snoRNA affected were indicative of the stress imposed. Increased temperature and downregulation of Nop58, a protein important to box C/D snoRNA stability, both expose structurally important stem structures. This reflects the molecular insights afforded by the systematic nature of the assay. It also reflects the uniqueness of each fitness landscape I derived despite many of the same positions being involved. The fitness landscapes at cold and high temperatures show major differences, the same as hypermorphic mutations in different genes do. Downregulation of Nop1 appears to alter the conformation of the pre-rRNA within SSU processome due to reduced methylation, lowering the tolerance for complimentary regions of U3 snoRNA to be mutated. This was contrary to the type of results I thought these experiments would find, which was mapping the positions important to binding of the protein downregulated. Instead the fitness landscape mirrors the selective pressures imposed. At low temperatures G-C

pairs are the most tolerant to mutation, while at higher temperatures they are the least. The results have also highlighted the influence of intragenic epistasis, with multiple mutations showing a greater magnitude of gene-gene and gene-environment interactions.

For strong fitness effects the assay can provide single nucleotide resolution, with position 31 showing heterogeneity depending of which substitution is present in the background of hypomorphic Nop1 levels. It should also be possible to use this data to look at individual substitutions or genotypes of lower effect, by first calculating noise as a function of coverage. This may also highlight genotypes with positive fitness effects. This along with testing other gene-gene interactions, such as other U3 snoRNA interacting proteins, or different environments presents avenues for future work.

References

Araya, C. & Fowler, D., 2011. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol,* 29(9), pp. 435-441.

Bank, C., Hietpas, R., Jensen, J. & Bolon, D., 2015. A systematic survey of an intragenic epistatic landscape. *Mol Biol Evol,* 32(1), pp. 229-238.

Barandun, J. et al., 2017. The complete structure of the small-subunit processome.. *Nat Struct Mol Biol,* Issue 24, pp. 944-953.

Barandun, J., Hunziker, M. & Klinge, S., 2018 . Assembly and structure of the SSU processome-a nucleolar precursor of the small ribosomal subunit. *Curr Opin Struct Biol,* pp. 85-93.

Baxter-Roshek, J., Petrov, A. & Dinman, J., 2007. Optimization of ribosome structure and function by rRNA base modification. *PLoS One.*

Beltrame, M. & Tollervey, D., 1995. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *EMBO J,* 14(17), pp. 4350-4356.

Bendixsen, D., Østman, B. & Hayden, E., 2017. Negative Epistasis in Experimental RNA Fitness Landscapes. *J Mol Evol,* 85(5-6), pp. 159-168.

Bershtein, S. et al., 2006. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature,* 444(7121), pp. 929-932.

Billy, E., Wegierski, T., Nasr, F. & Filipowicz, W., 2000 . Rcl1p, the yeast protein similar to the RNA 3'-phosphate cyclase, associates with U3 snoRNP and is required for 18S rRNA biogenesis. *EMBO ,* pp. 2115-26.

Bina-Stein, M. & Crothers, D., 1974. Conformational changes of transfer ribonucleic acid. The pH phase diagram under acidic conditions. *Biochemistry.*

Bisson, L., 1991. Influence of nitrogen on yeast and fermentation of grapes. *Proceedings of the International Symposium on Nitrogen in Grapes and Wine.*

Black, J., Wang, Z., Goering, L. & Johnson, A., 2018 . Utp14 interaction with the small subunit processome. *RNA,* 24(9), pp. 1214-1228.

Bleichert, F. et al., 2006 . The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. *Proc Natl Acad Sci U S A,* pp. 9464-9.

Borovjagin, A. & Gerbi, S., 2001. Xenopus U3 snoRNA GAC-Box A' and Box A sequences play distinct functional roles in rRNA processing. *Mol Cell Biol,* 21(18), pp. 6210-6221.

Boucher, B. & Jenna, S., 2013. Genetic interaction networks: better understand to better predict. *Front Genet.*

Boyle, E., Li, Y. & Pritchard, J., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell,* 169(7), pp. 1177-1186.

Breslow, D. et al., 2008. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods,* Issue 5, pp. 711-718.

Cahill, N. et al., 2002. Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J,* 21(14), pp. 3816-3828.

Chaker-Margot, M., 2018. Assembly of the small ribosomal subunit in yeast: mechanism and regulation. *RNA,* Issue 24.

Chaker-Margot, M., Barandun, J., Hunziker, M. & Klinge, S., 2017. Architecture of the yeast small subunit processome. *Science,* 355(6321).

Chaker-Margot, M. et al., 2015. Stage-specific assembly events of the 6-MDa small-subunit processome initiate eukaryotic ribosome biogenesis.. *Nat Struct Mol Biol,* Issue 22, pp. 920-923.

Cheng, J. et al., 2017. 3.2-Å-resolution structure of the 90S preribosome before A1 pre-rRNA cleavage. *Nat Struct Mol Biol,* 24(11), pp. 954-964.

Chursov, A. et al., 2012. Specific temperature-induced perturbations of secondary mRNA structures are associated with the cold-adapted temperature-sensitive phenotype of influenza A virus. *RNA Biol,* pp. 1266-1274.

Cléry, A. et al., 2007. Analysis of sequence and structural features that identify the B/C motif of U3 small nucleolar RNA as he recognition site for the Snu13p-Rrp9p protein pair.. *Mol Cell Biol.,* Issue 27, pp. 1191-1206.

Costanzo, M. et al., 2010. The genetic landscape of a cell. *Science,* 327(5964), pp. 425-431.

Daniels, M. et al., 1984. Cloning of genes involved in pathogenicity of Xanthomonas campestris pv. campestris using the broad host range cosmid pLAFR1. *EMBO J,* 3(13), pp. 3323-3328.

de Visser, J. & Krug, J., 2014. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet,* 15(7), pp. 480-490.

Delprato, A. et al., 2014 . Crucial role of the Rcl1p-Bms1p interaction for yeast pre-ribosomal RNA processing. *Nucleic Acids Res,* pp. 10161-72.

Deutschbauer, A. et al., 2005. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics,* 169(4), pp. 1915-1925.

Diss, G. & Lehner, B., 2018. The genetic landscape of a physical interaction. *Elife.*

Domingo, J., Diss, G. & Lehner, B., 2018. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature,* 558(7708), pp. 117-121.

Dragon, F. et al., 2002. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis.. *Nature,* Issue 417, pp. 967-970.

Dunbar, D., Wormsley, S., Agentis, T. & Baserga, S., 1997. Mpp10p, a U3 small nucleolar ribonucleoprotein component required for pre-18S rRNA processing in yeast.. *Mol Cell Biol,* Issue 17, pp. 5803-5812.

Dutca, L., Gallagher, J. & Baserga, S., 2011. The initial U3 snoRNA:pre-rRNA base pairing interaction required for pre-18S rRNA folding revealed by in vivo chemical probing.. *Nucleic Acids Res,* Issue 39, pp. 5164-5180.

Enquist, B. et al., 2003. Scaling metabolism from organisms to ecosystems. *Nature,* 423(6940), pp. 639-642.

Eyre-Walker, A. & Keightley, P., 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet,* 8(8), pp. 610-648.

Findlay, G. et al., 2014. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature,* 513(7516), pp. 120-123.

Flynn, K., Cooper, T., Moore, F. & Cooper, V., 2013. The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS Genet,* 9(4).

Freese EB, C. M. F. E., 1982. Initiation of yeast sporulation of partial carbon, nitrogen, or phosphate deprivation. *J Bacteriol,* 149(3), pp. 840-851.

Ganapathi, K. & Shimamura, A., 2008. Ribosomal dysfunction and inherited marrow failure. *Br J Haematol,* 141(3), pp. 376-387.

Ganot, P., Bortolin, M. & Kiss, T., 1997. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell,* 89(5), pp. 799-809.

Gasch, A. et al., 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell,* 11(12), pp. 4241-4257.

Gautier, T., Bergès, T., Tollervey, D. & Hurt, E., 1997. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol,* 17(12), pp. 7088-7098.

Gérczei, T. & Correll, C., 2004. Imp3p and Imp4p mediate formation of essential U3-precursor rRNA (pre-rRNA) duplexes, possibly to recruit the small subunit processome to the pre-rRNA. *Proc Natl Acad Sci U S A,* 101(43), pp. 15301-15306.

Gérczei, T. et al., 2009. RNA chaperones stimulate formation and yield of the U3 snoRNA-Pre-rRNA duplexes needed for eukaryotic ribosome biogenesis. *J Mol Biol,* 390(5), pp. 991-1006.

Giaever, G. et al., 2002. Functional profiling of the Saccharomyces cerevisiae genome. *Nature,* 418(6896), pp. 387-391.

Gietz, R. & Schiestl, R., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc.*

González, S., Barrio, E., Gafner, J. & Querol, A., 2006. Natural hybrids from Saccharomyces cerevisiae, Saccharomyces bayanus and Saccharomyces kudriavzevii in wine fermentations. *FEMS Yeast Res,* pp. 1221-1234.

Grandi, P. et al., 2002. 90S Pre-Ribosomes Include the 35S Pre-rRNA, the U3 snoRNP, and 40S Subunit Processing Factors but Predominantly Lack 60S Synthesis Factors. *Mol Cell,* Issue 10, pp. 105-115.

Granneman S, V. J. L. R. v. V. W. P. G. W. N., 2004. Role of pre-rRNA base pairing and 80S complex formation in subnucleolar localization of the U3 snoRNP.. *Mol Cell Biol,* Issue 24, pp. 4600-8610.

Granneman, S. et al., 2003. The human Imp3 and Imp4 proteins form a ternary complex with hMpp10, which only interacts with the U3 snoRNA in 60-80S ribonucleoprotein complexes. *Nucleic Acids Res,* 31(7), pp. 1877-1887.

Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D., 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs.. *Proc Natl Acad Sci,* Issue 106, pp. 9613-9618.

Granneman, S. et al., 2002. The hU3-55K protein requires 15.5K binding to the box B/C motif as well as flanking RNA elements for its association with the U3 small nucleolar RNA in Vitro.. *J Biol Chem,* Issue 277, pp. 48490-48500.

Hayden, E. & Wagner, A., 2012. Environmental change exposes beneficial epistatic interactions in a catalytic RNA. *Proc Biol Sci,* 279(1742), pp. 3418-3425.

Helm, M., 2006. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res,* 35(20), pp. 721-733.

Hemani, G. et al., 2014. Detection and replication of epistasis influencing transcription in humans. *Nature,* 508(7495), pp. 249-253.

Henras, A. et al., 2008. The post-transcriptional steps of eukaryotic ribosome biogenesis.. *Cell Mol Life Sci.,* Issue 65, pp. 2334-2359.

Hietpas, R., Jensen, J. & Bolon, D., 2011. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A,* 108(19), pp. 7896-7901.

Homma, T., Iwahashi, H. & Komatsu, Y., 2003. Yeast gene expression during growth at low temperature. *Cryobiology,* 46(3), pp. 230-237.

Horn, D., Mason, S. & Karbstein, K., 2011 . Rcl1 protein, a novel nuclease for 18S ribosomal RNA production. *J Biol Chem.*

Huang, H. & Brandriss, M., 2000. The regulator of the yeast proline utilization pathway is differentially phosphorylated in response to the quality of the nitrogen source. *Mol Cell Biol,* pp. 892-899.

Hughes, J. & Ares, M. J., 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5' external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J,* pp. 4231-4239.

Hunter, D., 2005. Gene-environment interactions in human diseases. *Nat Rev Genet,* 6(4), pp. 287-298.

Hunziker, M. et al., 2016. UtpA and UtpB chaperone nascent pre-ribosomal RNA and U3 snoRNA to initiate eukaryotic ribosome assembly.. *Nat Commun,* Issue 29.

Ide, S., Miyazaki, T., Maki, H. & Kobayashi, T., 2010. Abundance of ribosomal RNA gene copies maintains genome integrity.. *Science,* Issue 327, pp. 693-363.

Jazwinski, S., 1990. Aging and senescence of the budding yeast Saccharomyces cerevisiae. *Mol Microbiol,* 4(3), pp. 337-343.

Kass, S., Tyc, K., Steitz, J. & Sollner-Webb, B., 1990. The U3 Small Nucleolar Ribonucleuprotein Functions in the First Step of Preribosomal RNA Processing. *Cell,* Issue 60, pp. 897-908.

Kawai, S., Hashimoto, W. & Murata, K., 2010. Transformation of Saccharomyces cerevisiae and other fungi: methods and possible underlying mechanism. *Bioeng Bugs,* pp. 395-403.

Kiss-László, Z. et al., 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell,* 85(7), pp. 1077-1088.

Kiss-László, Z., Henry, Y. & Kiss, T., 1998. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J,* 17(3), pp. 797-807.

Kiss, T., 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J,* 20(14), pp. 3617-3622.

Klein, D., Schmeing, T., Moore, P. & Steitz, T., 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J,* 20(15), pp. 4214-4221.

Knox, A. et al., 2011. A weak C' box renders U3 snoRNA levels dependent on hU3-55K binding.. *Mol Cell Biol.,* Issue 31, pp. 2404-2412.

Kondo, K., Kowalski, L. & Inouye, M., 1992. Cold shock induction of yeast NSR1 protein and its role in pre-rRNA processing. *J Biol Chem,* pp. 16259-16265.

Kornprobst, M. et al., 2016 . Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell,* 166(2), pp. 389-393.

Kos-Braun, I., Jung, I. & Koš, M., 2017 . Tor1 and CK2 kinases control a switch between alternative ribosome biogenesis pathways in a growth-dependent manner. *PLoS Biol.*

Kos-Braun, I. & Koš, M., 2017. Post-transcriptional regulation of ribosome biogenesis in yeast. *Microb Cell,* 4(5), pp. 179-181.

Kos, M. & Tollervey, D., 2010. Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell,* pp. 809-820.

Kubota, T., Nishimura, K., Kanemaki, M. & Donaldson, A., 2013. The Elg1 replication factor C-like complex functions in PCNA unloading during DNA replication. *Mol Cell,* 50(2), pp. 273-280.

Kudla, G. et al., 2011. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci U S A,* 108(24), pp. 10010-10015.

Lafontaine, D. & Tollervey, D., 1999. Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA,* 5(3), pp. 455-467.

Lafontaine, D. & Tollervey, D., 2000. Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol Cell Biol,* 20(8), pp. 2650-2659.

Lee, J., Vogt, C., McBrairty, M. & Al-Hashimi, H., 2013. Influence of dimethylsulfoxide on RNA structure and ligand binding. *Anal Chem,* pp. 9692-9698.

Lee, S. & Baserga, S., 1997. Functional separation of pre-rRNA processing steps revealed by truncation of the U3 small nucleolar ribonucleoprotein component, Mpp10. *Proc Natl Acad Sci U S A,* 94(25), pp. 13536-13541.

Lee, S. & Baserga, S., 1999. Imp3p and Imp4p, two specific components of the U3 small nucleolar ribonucleoprotein that are essential for pre-18S rRNA processing. *Mol Cell Biol,* 19(8), pp. 5441-5452.

Lehner, B., 2011. Molecular mechanisms of epistasis within and between genes. *Trend Genet,* 27(8), pp. 323-331.

Liang, X., Liu, Q. & Fournier, M., 2009. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA. *RNA,* 15(9), pp. 1716-1728.

Li, C., Qian, W., Maclean, C. & Zhang, J., 2016. The fitness landscape of a tRNA gene. *Science,* 352(6287), pp. 837-840.

Li, C. & Zhang, J., 2018. Multi-environment fitness landscapes of a tRNA gene. *Nat Ecol Evol,* 2(6), pp. 1025-1032.

Lin, J. et al., 2011. Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature,* 469(7331), pp. 559-563.

Li, X. et al., 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity.. *Proc Natl Acad Sci.*

Lygerou, Z., Allmang, C., Tollervey, D. & Séraphin, B., 1996. Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science,* 272(5259), pp. 268-270.

Madhukar, N., Elemento, O. & Pandey, G., 2015. Prediction of Genetic Interactions Using Machine Learning and Network Properties. *Front Bioeng Biotechnol.*

Marmier-Gourrier, N. et al., 2011. A second base pair interaction between U3 small nucleolar RNA and the 5'-ETS region is required for early cleavage of the yeast pre-ribosomal RNA. *Nucleic Acids Res,* 39(22), pp. 9731-9745.

Martín, H. et al., 1993. Activity of the yeast MAP kinase homologue Slt2 is critically required for cell integrity at 37 degrees C. *Mol Gen Genet,* Volume 241, pp. 177-183.

Marz, M. & Stadler, P., 2009. Comparative analysis of eukaryotic U3 snoRNA. *RNA,* 6(5), pp. 503-507.

Matuszewski, S. et al., 2016. A Statistical Guide to the Design of Deep Mutational Scanning Experiments.. *Genetics,* Issue 204, pp. 77-87.

Melnikov, A. et al., 2014. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res,* 42(14).

Mendoza-Ochoa, G. et al., 2019 . A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. *Yeast,* pp. 75-81.

Méreau, A. et al., 1997. An in vivo and in vitro structure-function analysis of the Saccharomyces cerevisiae U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. *J Mol Biol,* 273(3), pp. 552-571.

Miller, O. J. & Beatty, B., 1969. Visualization of Nucleolar Genes.. *Science,* Issue 164, pp. 955-957.

Mohd Azhar, S. et al., 2017. Yeasts in sustainable bioethanol production: A review. *Biochem Biophys Rep.*

Moore, J., 2003. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered,* Issue 56, pp. 72-82.

Morawska, M. & Ulrich, H., 2013. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast,* 30(9), pp. 341-351.

Muhlrad, D. & Parker, R., 1999. Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA,* 5(10), pp. 1299-1307.

Myslinski, E., Ségault, V. & Branlant, C., 1990. An intron in the genes for U3 small nucleolar RNAs of the yeast Saccharomyces cerevisiae. *Science,* 247(4947), pp. 1213-1216.

Nabavi, S. & Nazar, R., 2008. U3 snoRNA promoter reflects the RNA's function in ribosome biogenesis. *Curr Genetics,* 54(4), pp. 175-184.

Narayanan, A., Speckmann, W., Terns, R. & Terns, M., 1999. Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli. *Mol Biol Cell,* Issue 10, pp. 2131-2147.

Nathan, D., Vos, M. & Lindquist, S., 1997. In vivo functions of the Saccharomyces cerevisiae Hsp90 chaperone. *Proc Natl Acad Sci U S A,* 94(24), pp. 12949-12956.

Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M., 2016. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep,* 15(1), pp. 210-218.

Neiman, A., 2011. Sporulation in the budding yeast Saccharomyces cerevisiae. *Genetics,* 189(3), pp. 737-765.

Nishimura, K. & Fukagawa, T., 2017. An efficient method to generate conditional knockout cell lines for essential genes by combination of auxin-inducible degron tag and CRISPR/Cas9. *Chromosome Res,* 25(3-4), pp. 253-260.

Nishimura, K. et al., 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods,* 6(12), pp. 917-922.

Nishimura, K. & Kanemaki, M., 2014. Rapid Depletion of Budding Yeast Proteins via the Fusion of an Auxin-Inducible Degron (AID). *Current Protocols in Cell Biology.*

Ochoa, G., 2017. Using the auxin-inducible degron to study the spliceosome cycle and splicing fidelity. *University of Edinburgh thesis.*

Ortiz-Tovar G, P.-T. R. A. A. B. E. Q. A., 2018. A comparison of the performance of natural hybrids Saccharomyces cerevisiae × Saccharomyces kudriavzevii at low temperatures reveals the crucial role of their S. kudriavzevii genomic contribution. *Int J Food Microbiol,* pp. 12-19.

Osheim, Y. et al., 2004. Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. *Mol Cell,* 16(6), pp. 943-954.

Ottman, R., 1996. Gene-environment interaction: definitions and study designs. *Prev Med,* 25(6), pp. 764-770.

Palmer, A. et al., 2015. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat Commun.*

Papagiannakis, A., de Jonge, J., Zhang, Z. & Heinemann, M., 2017. Quantitative characterization of the auxin-inducible degron: a guide for dynamic protein depletion in single yeast cells. *Sci Rep,* 7(1).

Pérez-Fernández, J. et al., 2007 . The 90S preribosome is a multimodular structure that is assembled through a hierarchical mechanism. *Mol Cell Biol,* 27(15), pp. 5414-5429.

Phillips, P., 1998. The language of gene interaction. *Genetics,* 149(3), pp. 1167-1171.

Phipps, K., Charette, J. & Baserga, S., 2010. The small subunit processome in ribosome biogenesis—progress and prospects. *Wiley Interdiscip Rev RNA,* Issue 2, pp. 1-21.

Pitt, J. & Ferré-D'Amaré, A., 2010. Rapid construction of empirical RNA fitness landscapes. *Science,* 330(6002), pp. 376-379.

Poelwijk, F., Kiviet, D., Weinreich, D. & Tans, S., 2007. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature,* 445(7126), pp. 383-386.

Pretorius, I., 2000. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast,* pp. 675-729.

Puchta, O. et al., 2016. Network of epistatic interactions within a yeast snoRNA. *Science,* Issue 352, pp. 840-844.

Qian, W. et al., 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep,* pp. 1399-1410.

Rollins, N. et al., 2018. 3D protein structure from genetic epistasis experiments. *biorxiv.*

Rowley, A. et al., 1993. Heat shock-mediated cell cycle blockage and G1 cyclin expression in the yeast Saccharomyces cerevisiae. *Mol Cell Biol,* 13(2), pp. 1034-1041.

Ruggero, D. & Pandolfi, P., 2003. Does the ribosome translate cancer?. *Nat Rev Cancer.,* Issue 3, pp. 179-192.

Sahara, T., Goda, T. & Ohgiya, S., 2002. Comprehensive expression analysis of time-dependent genetic responses in yeast cells to low temperature. *J Biol Chem,* 277(51), pp. 50015-50021.

Samarsky, D. & Fournier, M., 1998. Functional mapping of the U3 small nucleolar RNA from the yeast Saccharomyces cerevisiae.. *Mol Cell Biol,* Issue 18, pp. 3431-3444.

Sanchez, Y. & Lindquist, S., 1990. HSP104 required for induced thermotolerance. *Science,* 248(4959), pp. 1112-1115.

Sanjuán, R., Moya, A. & Elena, S., 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci U S A,* 101(22), pp. 8396-8401.

Sardana, R. et al., 2015. The DEAH-box helicase Dhr1 dissociates U3 from the pre-rRNA to promote formation of the central pseudoknot. *PLoS Biol,* Issue 13.

Sarkisyan, K. et al., 2016. Local fitness landscape of the green fluorescent protein. *Nature,* 533(7603), pp. 397-401.

Schimmang, T. et al., 1989. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J,* 8(13), pp. 4015-4024.

Schmiedel, J. & Lehner, B., 2018. Determining protein structures using genetics. *biorxiv.*

Shah, B., Liu, X. & Correll, C., 2013. Imp3 unfolds stem structures in pre-rRNA and U3 snoRNA to form a duplex essential for small subunit processing. *RNA,* 19(10), pp. 1372-1383.

Sharma, K. & Tollervey, D., 1999. Base pairing between U3 small nucleolar RNA and the 5' end of 18S rRNA is required for pre-rRNA processing. *Mol Cell Biol,* 19(9), pp. 6012-6019.

Sjöblom, T. et al., 2006. The consensus coding sequences of human breast and colorectal cancers.. *Science,* Issue 314, pp. 268-274.

Sondalle, S. & Baserga, S., 2014. Human diseases of the SSU processome. *Biochim Biophys Acta,* Issue 1842, pp. 758-764.

Speckmann, W., Narayanan, A., Terns, R. & Terns, M., 1999. Nuclear retention elements of U3 small nucleolar RNA. *Mol Cell Biol,* 19(12), pp. 8412-8421.

Speckmann, W., Terns, R. & Terns, M., 2000. The box C/D motif directs snoRNA 5'-cap hypermethylation.. *Nucleic Acids Res.,* Issue 28, pp. 4467-4473.

Steinberg, B. & Ostermeier, M., 2016. Environmental changes bridge evolutionary valleys. *Sci Adv.*

Strauss, E. & Guthrie, C., 1991. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Gene Dev,* 5(4), pp. 629-641.

Su, H. et al., 2014. Elevated snoRNA biogenesis is essential in breast cancer.. *Oncogene,* Issue 33, pp. 1348-1358.

Sun, Q. et al., 2017. Molecular architecture of the 90S small subunit pre-ribosome. *Elife.*

Szewczak, L. et al., 2005. Molecular basis for RNA kink-turn recognition by the h15.5K small RNP protein. *RNA,* 11(9), pp. 1407-1419.

Tanaka, S., Miyazawa-Onami, M., Iida, T. & Araki, H., 2015. iAID: an improved auxin-inducible degron system for the construction of a 'tight'

conditional mutant in the budding yeast Saccharomyces cerevisiae. *Yeast,* 32(8), pp. 567-581.

Tan, R. & van Oudenaarden, A., 2010. Transcript counting in single cells reveals dynamics of rDNA transcription.. *Mol Syst Biol,* Issue 358.

Thiry, M. & Lafontaine, D., 2005. Birth of a nucleolus: the evolution of nucleolar compartments. *Trends Cell Biol,* Issue 15, pp. 194-199.

Tollervey, D., Lehtonen, H., Carmo-Fonseca, M. & Hurt, E., 1991. The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. *EMBO J,* 10(3), pp. 573-583.

Tollervey, D. et al., 1993. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell,* 72(3), pp. 443-457.

Tong, A. et al., 2004. Global mapping of the yeast genetic interaction network. *Science,* 303(5659), pp. 808-813.

Tronchoni, J. et al., 2014. Transcriptomics of cryophilic Saccharomyces kudriavzevii reveals the key role of gene translation efficiency in cold stress adaptations. *BMC Genomics.*

Trost, M., Blattner, A. & Lehner, C., 2016. Regulated protein depletion by the auxin-inducible degradation system in Drosophila melanogaster. *Fly,* 10(1), pp. 35-46.

Udem, S. & Warner, J., 1972. Ribosomal RNA synthesis in Saccharomyces cerevisiae. *J Mol Biol,* 65(2), pp. 227-242.

Venema, J. et al., 2000. Yeast Rrp9p is an evolutionarily conserved U3 snoRNP protein essential for early pre-rRNA processing cleavages and requires box C for its association.. *RNA,* Issue 6, pp. 1660-1671.

Verheggen, C. et al., 2002 . Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J,* 21(11), pp. 2736-2745.

Vincent, N., Charette, J. & Baserga, S., 2018. The SSU processome interactome in Saccharomyces cerevisiae reveals novel protein subcomplexes. *RNA,* 24(1), pp. 77-89.

von der Haar, T., 2008. A quantitative estimation of the global translational activity in logarithmically growing yeast cells.. *BMC Syst Biol,* Issue 2.

Wan, Y. et al., 2012. Genome-wide measurement of RNA folding energies. *Moll Cell,* 48(2), pp. 169-181.

Warner, J., 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci,* Issue 24, pp. 437-440.

Watase, G., Takisawa, H. & Kanemaki, M., 2012. Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS. *Curr Biol,* 22(4), pp. 343-349.

Watkins, N. & Bohnsack, M., 2012. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA.. *Wiley Interdiscip Rev RNA,* Issue 3, pp. 397-414.

Watkins, N. et al., 2004. Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol Cell,* 16(5), pp. 789-798.

Watkins, N. et al., 2000. A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP.. *Cell,* Issue 103, pp. 457-466.

Weinberg, R. A. & Penman, S., 1968. Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.,* pp. 289-304.

Weinreich, D., Delaney, N., Depristo, M. & Hartl, D., 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science,* 312(5770), pp. 111-114.

Weinreich, D., Lan, Y., Jaffe, J. & Heckendorn, R., 2018. The Influence of Higher-Order Epistasis on Biological Fitness Landscape Topography. *J Stat Phys,* 172(1), pp. 208-225.

Wells, G. et al., 2016. The PIN domain endonuclease Utp24 cleaves pre-ribosomal RNA at two coupled sites in yeast and humans. *Nucleic Acids Res,* 44(18), pp. 5399-5409.

Woolford, J. & Baserga, S., 2013. Ribosome biogenesis in the yeast Saccharomyces cerevisiae.. *Genetics,* Issue 195, pp. 643-681.

Wormsley, S., Samarsky, D., Fournier, M. & Baserga, S., 2001. An unexpected, conserved element of the U3 snoRNA is required for Mpp10p association. *RNA,* 7(6), pp. 904-919.

Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth Internation Congress of Geneics 1,* pp. 356-366.

Wu, M. & Ma, S., 2018. Robust genetic interaction analysis. *Brief Bioinform.*

Zavanelli, M., Britton, J., Igel, A. & Ares, M. J., 1994. Mutations in an essential U2 small nuclear RNA structure cause cold-sensitive U2 small

nuclear ribonucleoprotein function by favoring competing alternative U2 RNA structures. *Mol Cell Biol,* pp. 1689-1697.

Zebarjadian, Y. et al., 1999. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol Cell Biol,* 19(11), pp. 7461-7472.

Zhang L, L. J. Y. K., 2013. Structural and functional analysis of the U3 snoRNA binding protein Rrp9. *RNA,* Issue 19, pp. 701-711.

Zhang, L., Ward, J., Cheng, Z. & Dernburg, A., 2015. The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans.. *Development,* 142(24), pp. 4374-4384.

Zhang, L. et al., 2016 . Stepwise and dynamic assembly of the earliest precursors of small ribosomal subunits in yeast. *Genes Dev,* 30(6), pp. 718-732.